

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:37:44 ; Search time 15.4854 Seconds  
(without alignments)  
1829.738 Million cell updates/sec

Title: US-10-073-333A-2  
Perfect score: 963  
Sequence: 1 MLFRARGPVRGWRGWRPAEA.....TCYFWKKKQKKKTKILNLFN 963

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	0.9	1487	2 US-08-760-489-2	Sequence 2, Appli
2	9	0.9	1487	2 US-08-760-489-4	Sequence 4, Appli
3	9	0.9	1487	4 US-09-185-373-2	Sequence 2, Appli
4	9	0.9	1487	4 US-09-185-373-4	Sequence 4, Appli
5	8	0.8	100	1 US-08-158-189-7	Sequence 7, Appli
6	8	0.8	133	4 US-08-468-560C-8	Sequence 8, Appli
7	8	0.8	139	2 US-08-219-237B-8	Sequence 8, Appli
8	8	0.8	140	4 US-08-477-347-17	Sequence 17, Appli
9	8	0.8	140	4 US-08-476-862-8	Sequence 8, Appli
10	8	0.8	153	4 US-09-286-529-2	Sequence 2, Appli
11	8	0.8	205	3 US-08-974-022-51	Sequence 51, Appli
12	8	0.8	205	4 US-08-795-445A-51	Sequence 51, Appli
13	8	0.8	205	4 US-08-795-447A-51	Sequence 51, Appli
14	8	0.8	205	4 US-08-974-186-51	Sequence 51, Appli
15	8	0.8	205	4 US-08-795-446B-51	Sequence 51, Appli
16	8	0.8	205	4 US-08-706-945D-138	Sequence 138, App
17	8	0.8	206	1 US-08-097-827-7	Sequence 7, Appli
18	8	0.8	206	1 US-08-494-574-7	Sequence 7, Appli
19	8	0.8	211	4 US-09-286-529-20	Sequence 20, Appli
20	8	0.8	299	4 US-09-286-529-17	Sequence 17, Appli
21	8	0.8	300	2 US-08-794-796-2	Sequence 2, Appli
22	8	0.8	438	1 US-08-097-827-11	Sequence 11, Appli
23	8	0.8	438	1 US-08-494-574-11	Sequence 11, Appli
24	7	0.7	15	4 US-08-602-999A-363	Sequence 363, App
25	7	0.7	15	4 US-09-500-124-363	Sequence 363, App
26	7	0.7	77	4 US-09-146-950-25	Sequence 25, Appli
27	7	0.7	155	4 US-09-146-950-4	Sequence 4, Appli

28	7	0.7	159	4	US-09-146-950-20	Sequence 20, Appli
29	7	0.7	193	4	US-09-146-950-2	Sequence 2, Appli
30	7	0.7	197	4	US-09-146-950-18	Sequence 18, Appli
31	7	0.7	278	2	US-08-569-168-2	Sequence 2, Appli
32	7	0.7	283	4	US-08-509-024-2	Sequence 2, Appli
33	7	0.7	283	4	US-09-333-279-2	Sequence 2, Appli
34	7	0.7	283	4	US-09-072-993C-2	Sequence 2, Appli
35	7	0.7	283	5	PCT-US96-12374-2	Sequence 2, Appli
36	7	0.7	320	2	US-08-933-750C-12	Sequence 12, Appli
37	7	0.7	320	4	US-09-234-613-12	Sequence 12, Appli
38	7	0.7	331	4	US-09-086-483A-3	Sequence 3, Appli
39	7	0.7	383	2	US-08-569-168-7	Sequence 7, Appli
40	7	0.7	419	4	US-08-509-024-7	Sequence 7, Appli
41	7	0.7	419	4	US-09-333-279-7	Sequence 7, Appli
42	7	0.7	479	4	US-09-177-349-3	Sequence 3, Appli
43	7	0.7	588	5	PCT-US95-13749-4	Sequence 4, Appli
44	7	0.7	761	1	US-07-906-395-2	Sequence 2, Appli
45	7	0.7	761	1	US-08-192-632-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-760-489-2  
; Sequence 2, Application US/08760489  
; Patent No. 5830696  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay M.  
; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC  
; ENZYMES  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/760,489  
; FILING DATE: 05-DEC-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/008,311  
; FILING DATE: 07-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haile, Ph.D., Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 09010/008001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5099  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1487 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-08-760-489-2

Query Match 0.9%; Score 9; DB 2; Length 1487;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 EAPRRGRSP 27  
|||||||



Db 171 EAPRRGRSP 179

RESULT 2

US-08-760-489-4

; Sequence 4, Application US/08760489

; Patent No. 5830696

; GENERAL INFORMATION:

; APPLICANT: Short, Jay M.

; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC

; ENZYMES

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson, P.C.

; STREET: 4225 Executive Square, Suite 1400

; CITY: La Jolla

; STATE: CA

; COUNTRY: US

; ZIP: 92037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows95

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/760,489

; FILING DATE: 05-DEC-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/008,311

; FILING DATE: 07-DEC-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Haile, Ph.D., Lisa A.

; REGISTRATION NUMBER: 38,347

; REFERENCE/DOCKET NUMBER: 09010/008001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 619/678-5070

; TELEFAX: 619/678-5099

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1487 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FRAGMENT TYPE: internal

US-08-760-489-4

Query Match 0.9%; Score 9; DB 2; Length 1487;

Best Local Similarity 100.0%; Pred. No. 3.6;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 EAPRRGRSP 27

|||||

Db 171 EAPRRGRSP 179

RESULT 3

US-09-185-373-2

; Sequence 2, Application US/09185373

; Patent No. 6335179

; GENERAL INFORMATION:

; APPLICANT: Short, Jay M.

; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC

; ENZYMES

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson, P.C.

; STREET: 4225 Executive Square, Suite 1400

; CITY: La Jolla

; STATE: CA

; COUNTRY: US

; ZIP: 92037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows95

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/185,373

; FILING DATE: 03-No. 6335179-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/760,489

; FILING DATE: 05-DEC-1996

; APPLICATION NUMBER: 60/008,311

; FILING DATE: 07-DEC-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Haile, Ph.D., Lisa A.

; REGISTRATION NUMBER: 38,347

; REFERENCE/DOCKET NUMBER: 09010/008001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 619/678-5070

; TELEFAX: 619/678-5099

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1487 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FRAGMENT TYPE: internal

; SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-185-373-2

Query Match 0.9%; Score 9; DB 4; Length 1487;

Best Local Similarity 100.0%; Pred. No. 3.6;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 EAPRRGRSP 27

|||||

Db 171 EAPRRGRSP 179

RESULT 4

US-09-185-373-4

; Sequence 4, Application US/09185373

; Patent No. 6335179

; GENERAL INFORMATION:

; APPLICANT: Short, Jay M.

; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC

; ENZYMES

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson, P.C.

; STREET: 4225 Executive Square, Suite 1400

; CITY: La Jolla

; STATE: CA

; COUNTRY: US

; ZIP: 92037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows95

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/185,373

; FILING DATE: 03-No. 6335179-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/760,489

; FILING DATE: 05-DEC-1996

; APPLICATION NUMBER: 60/008,311

; FILING DATE: 07-DEC-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Haile, Ph.D., Lisa A.

; REGISTRATION NUMBER: 38,347

; REFERENCE/DOCKET NUMBER: 09010/008001

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1487 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-185-373-4
Query Match 0.9%; Score 9; DB 4; Length 1487;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 EAPRRGRSP 27
Db 171 EAPRRGRSP 179

RESULT 5
US-08-158-189-7
; Sequence 7, Application US/08158189
; Patent No. 5641497
; GENERAL INFORMATION:
; APPLICANT: Bevins, Charles L.
; APPLICANT: Jones, Douglas E.
; TITLE OF INVENTION: Gastrointestinal Defensin Peptides,
; TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/158,189
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,232
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Johnson, Philip S.
; REGISTRATION NUMBER: 27,200
; REFERENCE/DOCKET NUMBER: CH-0219
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-158-189-7
Query Match 0.8%; Score 8; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 936 TAVLLVAL 943
Db 8 TAVLLVAL 15
```

```

RESULT 6
US-08-468-560C-8
; Sequence 8, Application US/08468560C
; Patent No. 6270998
; GENERAL INFORMATION:
; APPLICANT: NAGATA, Shigekazu
; APPLICANT: ITOH, Naoto
; APPLICANT: YONEHARA, Shin
; TITLE OF INVENTION: DNA CODING FOR HUMAN CELL SURFACE
; TITLE OF INVENTION: ANTIGEN
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP.
; STREET: P.O. BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,560C
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR., GERLAD M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 20-4393P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 133 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-468-560C-8
Query Match 0.8%; Score 8; DB 4; Length 133;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 616 CVPCPPGH 623
Db 91 CVPCPPGH 98

RESULT 7
US-08-219-237B-8
; Sequence 8, Application US/08219237B
; Patent No. 5874546
; GENERAL INFORMATION:
; APPLICANT: NAGATA, Shigekazu
; APPLICANT: ITOH, Naoto
; APPLICANT: YONEHARA, Shin
; TITLE OF INVENTION: DNA Coding for Human Cell Surface Antigen
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James W. Hellwege
; STREET: P.O. Box 2266 Eads Station
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/219,237B  
FILING DATE: 28-MAR-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/872,129  
FILING DATE: 22-APR-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: James W. Hellwege  
REGISTRATION NUMBER: 28,808  
REFERENCE/DOCKET NUMBER: 516762  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 139 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-219-237B-8

Query Match 0.8%; Score 8; DB 2; Length 139;  
Best Local Similarity 100.0%; Pred. No. 4.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPGGH 623  
|||||||  
Db 97 CVPCPGGH 104

## RESULT 8

US-08-477-347-17  
Sequence 17, Application US/08477347  
Patent No. 6232446  
GENERAL INFORMATION:  
APPLICANT: WALLACH, David  
APPLICANT: BIGDA, Jacek  
APPLICANT: BELETSKY, Igor  
APPLICANT: METT, Igor  
TITLE OF INVENTION: TNF LIGANDS  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,347  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/115,685  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 106271  
FILING DATE: 08-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Townsend, G. Kevin  
REGISTRATION NUMBER: 34,033  
REFERENCE/DOCKET NUMBER: WALLACH-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633

INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 140 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-477-347-17

Query Match 0.8%; Score 8; DB 4; Length 140;  
Best Local Similarity 100.0%; Pred. No. 4.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPGGH 623  
|||||||  
Db 98 CVPCPGGH 105

## RESULT 9

US-08-476-862-8  
Sequence 8, Application US/08476862  
Patent No. 6262239  
GENERAL INFORMATION:  
APPLICANT: WALLACH, David  
APPLICANT: BIGDA, Jacek  
APPLICANT: BELETSKY, Igor  
APPLICANT: METT, Igor  
APPLICANT: ENGELMANN, Hartmut  
TITLE OF INVENTION: TNF INHIBITORS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/476,862  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 107267  
FILING DATE: 12-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 94039  
FILING DATE: 06-APR-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 91229  
FILING DATE: 06-AUG-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 90339  
FILING DATE: 18-MAY-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: WALLACH-12A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 140 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-476-862-8

Query Match 0.8%; Score 8; DB 4; Length 140;  
Best Local Similarity 100.0%; Pred. No. 4.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623  
Db 98 CVPCPPGH 105

RESULT 10  
US-09-286-529-2  
; Sequence 2, Application US/09286529  
; Patent No. 6297367  
; GENERAL INFORMATION:  
; APPLICANT: Catherine Tribouley  
; TITLE OF INVENTION: NEW MEMBERS OF TNF AND TNFR FAMILIES  
; FILE REFERENCE: 1408.003/200130.439C1  
; CURRENT APPLICATION NUMBER: US/09/286,529  
; CURRENT FILING DATE: 1999-04-05  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 153  
; TYPE: PRT  
; ORGANISM: human  
US-09-286-529-2

Query Match 0.8%; Score 8; DB 4; Length 153;  
Best Local Similarity 100.0%; Pred. No. 4.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 PCPPGTFS 417  
Db 68 PCPPGTFS 75

RESULT 11  
US-08-974-022-51  
; Sequence 51, Application US/08974022  
; Patent No. 6015938  
; GENERAL INFORMATION:  
; APPLICANT: Boyle, William J.  
; APPLICANT: Lacey, David L.  
; APPLICANT: Calzone, Frank J.  
; APPLICANT: Chang, Ming-Shi  
; TITLE OF INVENTION: OSTEOPROTEGERIN  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: 1840 Dehavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/974,022  
; FILING DATE: 12-DEC-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/577,788  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.  
; REFERENCE/DOCKET NUMBER: A-378  
; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 205 amino acids

; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-974-022-51

Query Match 0.8%; Score 8; DB 3; Length 205;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623  
Db 122 CVPCPPGH 129

RESULT 12  
US-08-795-445A-51  
; Sequence 51, Application US/08795445A  
; Patent No. 6284485  
; GENERAL INFORMATION:  
; APPLICANT: Boyle, William J.  
; APPLICANT: Lacey, David L.  
; APPLICANT: Calzone, Frank J.  
; APPLICANT: Chang, Ming-Shi  
; TITLE OF INVENTION: OSTEOPROTEGERIN  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: 1840 Dehavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/795,445A  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/577,788  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.  
; REFERENCE/DOCKET NUMBER: A-378  
; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 205 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-795-445A-51

Query Match 0.8%; Score 8; DB 4; Length 205;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623  
Db 122 CVPCPPGH 129

RESULT 13  
US-08-795-447A-51  
; Sequence 51, Application US/08795447A  
; Patent No. 6284728  
; GENERAL INFORMATION:  
; APPLICANT: Boyle, William J.  
; APPLICANT: Lacey, David L.



APPLICANT: Calzone, Frank J.  
APPLICANT: Chang, Ming-Shi  
TITLE OF INVENTION: Osteoprotegerin  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: One Amgen Center Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91362-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/795,447A  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-378D2  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 205 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-795-447A-51

Query Match 0.8%; Score 8; DB 4; Length 205;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623  
Db 122 CVPCPPGH 129

RESULT 14  
US-08-974-186-51  
Sequence 51, Application US/08974186  
Patent No. 6284740  
GENERAL INFORMATION:  
APPLICANT: Boyle, William J.  
APPLICANT: Lacey, David L.  
APPLICANT: Calzone, Frank J.  
APPLICANT: Chang, Ming-Shi  
TITLE OF INVENTION: OSTEOPROTEGERIN  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: 1840 Dehavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,186  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/577,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.

REFERENCE/DOCKET NUMBER: A-378  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 205 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-974-186-51

Query Match 0.8%; Score 8; DB 4; Length 205;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623  
Db 122 CVPCPPGH 129

RESULT 15  
US-08-795-446B-51  
Sequence 51, Application US/08795446B  
Patent No. 6288032  
GENERAL INFORMATION:  
APPLICANT: Boyle, William J.  
APPLICANT: Lacey, David L.  
APPLICANT: Calzone, Frank J.  
APPLICANT: Chang, Ming-Shi  
TITLE OF INVENTION: OSTEOPROTEGERIN  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: 1840 Dehavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/795,446B  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/577,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-378  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 205 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-795-446B-51

Query Match 0.8%; Score 8; DB 4; Length 205;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623  
Db 122 CVPCPPGH 129

Search completed: May 12, 2003, 13:42:22  
Job time : 17.4854 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.  
OM protein - protein search, using sw model  
Run on: May 12, 2003, 13:30:09 ; Search time 31.4548 Seconds  
(without alignments)  
6308.212 Million cell updates/sec  
Title: US-10-073-333A-2  
Perfect score: 963  
Sequence: 1 MLFRARGPVGRGWGRPAEA.....TCYFWKKNQKKKKTILNLFN 963  
Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0  
Searched: 671580 seqs, 206047115 residues  
Word size : 0  
Total number of hits satisfying chosen parameters: 671580  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Listing first 45 summaries  
Database : SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length	DB ID	Description
1	342	35.5	493	4	Q96DP2
2	11	1.1	300	11	Q8R215
3	11	1.1	580	4	Q9P2M2
4	9	0.9	552	10	Q82444
5	8	0.8	55	12	Q67608
6	8	0.8	148	17	Q9HR77
7	8	0.8	154	12	Q9DGZ5
8	8	0.8	154	12	Q9E6T5
9	8	0.8	239	16	Q9ABT4
10	8	0.8	268	16	Q926W3
11	8	0.8	268	16	Q8Y3P0
12	8	0.8	281	5	Q9TZN3
13	8	0.8	286	16	Q9PR14
14	8	0.8	293	12	Q8V9I1
15	8	0.8	379	2	Q50198
16	8	0.8	380	8	Q9TFR5

17	8	0.8	380	8	Q958X6
18	8	0.8	428	16	Q8ZFL4
19	8	0.8	444	10	O23079
20	8	0.8	447	5	O44593
21	8	0.8	498	4	O43267
22	8	0.8	550	4	Q9P2Q1
23	8	0.8	593	10	Q8RUU6
24	8	0.8	597	10	Q9LGL5
25	8	0.8	605	10	Q8S4W8
26	8	0.8	606	10	Q8S4W9
27	8	0.8	607	10	Q9FFT4
28	8	0.8	607	10	Q96536
29	8	0.8	622	2	Q45155
30	8	0.8	704	10	O04434
31	8	0.8	721	11	Q922S8
32	8	0.8	742	16	Q8YNZ6
33	8	0.8	881	10	O65924
34	8	0.8	1051	12	Q91EU4
35	8	0.8	1059	10	Q9LHS4
36	8	0.8	1206	16	Q9CCX9
37	8	0.8	1232	10	Q9LJQ1
38	8	0.8	1578	16	Q92E25
39	8	0.8	1582	16	Q8Y9A5
40	8	0.8	2295	5	Q9TY98
41	8	0.8	2712	16	Q9F3X5
42	8	0.8	3102	5	Q9TZR4
43	8	0.8	3503	5	Q24292
44	7	0.7	65	6	Q95J37
45	7	0.7	79	2	Q9AFQ6

ALIGNMENTS

RESULT 1  
Q96DP2  
ID Q96DP2 PRELIMINARY; PRT; 493 AA.  
AC Q96DP2;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE CDNA FLJ31340 fis, clone MESAN1000035, weakly similar to major surface-labeled trophozoite antigen precursor.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK055902; BAB71041.1; -  
DR InterPro; IPR001881; EGF\_Ca.  
DR SMART; SM00179; EGF\_CA; 1.  
SQ SEQUENCE 493 AA; 54585 MW; 0AEBB2C19709B211 CRC64;

Query Match 35.5%; Score 342; DB 4; Length 493;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	334	GSSECTERPPCTTKDYFQIHPTCDEEGKTOIMYKWIPIKICREDLTDAIRLPPSGEKKDC	393
Db	96	GSSECTERPPCTTKDYFQIHPTCDEEGKTOIMYKWIPIKICREDLTDAIRLPPSGEKKDC	155
QY	394	PPCNPGFYNNGSSSCHPCPPPGTFFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCF	453

Db 156 PPCNPGFYNNSSSCHPCPPGTFSDGCKEPCPCAGTEPALGFYKWNVLPGNMKTSF 215  
QY 454 NVGNSKCDGMNGWEVAGDHQSGAGGSDNDYLILNLHIPGFKPPTSMGTGATGSELGRITF 513  
Db 216 NVGNSKCDGMNGWEVAGDHQSGAGGSDNDYLILNLHIPGFKPPTSMGTGATGSELGRITF 275  
QY 514 VFETLCSADCVLYFMVDINRKSTNVVSWGGTKEKQAYTHIIFKNATFTTWFQRTNOG 573  
Db 276 VFETLCSADCVLYFMVDINRKSTNVVSWGGTKEKQAYTHIIFKNATFTTWFQRTNOG 335  
QY 574 QDNRRFINDMVKIYSITATNAVGVASSCRACALGSEQSGSSCVPCPPGHVIEKETNQCK 633  
Db 336 QDNRRFINDMVKIYSITATNAVGVASSCRACALGSEQSGSSCVPCPPGHVIEKETNQCK 395  
QY 634 ECPPDITYLSIHQVYGEACIPCGPGSKNNQDHSVCYSDCFFY 675  
Db 396 ECPPDITYLSIHQVYGEACIPCGPGSKNNQDHSVCYSDCFFY 437  
RESULT 2  
Q8R215  
ID Q8R215 PRELIMINARY; PRT; 300 AA.  
AC Q8R215;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical 32.6 kDa protein (Fragment).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=COLON;  
RA Strausberg R.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC022655; AAH22655.1; -.  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 300 AA; 32642 MW; 6455109C054B6CB4 CRC64;  
Query Match 1.1%; Score 11; DB 11; Length 300;  
Best Local Similarity 100.0%; Pred. No. 0.015;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 943 LTCYFWKKKQK 953  
Db 213 LTCYFWKKKQK 223  
RESULT 3  
Q9P2M2  
ID Q9P2M2 PRELIMINARY; PRT; 580 AA.  
AC Q9P2M2;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE KIAA1324 protein (Fragment).  
GN KIAA1324.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=20181126; PubMed=10718198;  
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes.XVI.  
RT The complete sequences of 150 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 7:65-73(2000).

DR EMBL; AB037745; BAA92562.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 580 AA; 63365 MW; 5E63C19265EC8E5C CRC64;  
Query Match 1.1%; Score 11; DB 4; Length 580;  
Best Local Similarity 100.0%; Pred. No. 0.027;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 510 RITFVFETLCS 520  
Db 72 RITFVFETLCS 82  
RESULT 4  
O82444  
ID O82444 PRELIMINARY; PRT; 552 AA.  
AC O82444;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Peroxisomal targeting sequence 1 receptor (Fragment).  
GN PEX5.  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99007315; PubMed=9789089;  
RA Kragler F., Lametschwandtner G., Christmann J., Hartig A.,  
RA Harada J.J.;  
RT "Identification and analysis of the plant peroxisomal targeting signal  
RT 1 receptor NtPEX5.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:13336-13341(1998).  
DR EMBL; AF056282; AAC69180.1; -.  
DR InterPro; IPR001440; TPR.  
DR Pfam; PF00515; TPR; 4.  
DR SMART; SM00028; TPR; 4.  
KW Receptor.  
FT NON\_TER 1  
SQ SEQUENCE 552 AA; 62293 MW; 7045FA177B0F51C6 CRC64;  
Query Match 0.9%; Score 9; DB 10; Length 552;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 RARGPVRGR 12  
Db 22 RARGPVRGR 30  
RESULT 5  
Q67608  
ID Q67608 PRELIMINARY; PRT; 55 AA.  
AC Q67608;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Tomato golden mosaic virus subgenomic DNA derived from DNA B cccds -  
DE covalently closed circular double-stranded molecule.  
OS Tomato golden mosaic virus (TGMV).  
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.  
OX NCBI\_TaxID=10831;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87040767; PubMed=3022243;  
RA MacDowell S.W., Coutts R.H.A., Buck K.W.;  
RT "Molecular characterization of subgenomic single-stranded and double-  
RT stranded DNA forms isolated from plants infected with tomato golden  
RT mosaic virus.";  
RL Nucleic Acids Res. 14:7967-7984(1986).  
DR EMBL; X04485; CAA28171.1; -.

DR InterPro: IPR000211; Gemini\_BL.  
DR Pfam: PF00845; Gemini\_BL1; 1.  
SQ SEQUENCE 55 AA; 6281 MW; 41963570D739EC96 CRC64;

Query Match 0.8%; Score 8; DB 12; Length 55;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 NYIESNRD 184  
Db 13 NYIESNRD 20

RESULT 6  
Q9HR77 PRELIMINARY; PRT; 148 AA.  
AC Q9HR77;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Vng0825c.  
GN VNG0825C.  
OS Halobacterium sp. (strain NRC-1).  
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
OC Halobacteriaceae; Halobacterium.  
OX NCBI\_TaxID=64091;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20504483; PubMed=11016950;  
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,  
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,  
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;  
RT "Genome sequence of Halobacterium species NRC-1.";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
DR EMBL; AE005024; AAG19281.1; -  
DR InterPro: IPR004435; MobB.  
DR Pfam; PF03205; MobB; 1.  
KW Complete proteome.  
SQ SEQUENCE 148 AA; 15073 MW; F242336129C362FE CRC64;

Query Match 0.8%; Score 8; DB 17; Length 148;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 TVVGPSSDS 162  
Db 5 TVVGPSSDS 12

RESULT 7  
Q9DGZ5 PRELIMINARY; PRT; 154 AA.  
AC Q9DGZ5;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE X protein.  
GN X.  
OS Hepatitis B virus.  
OC Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.  
OX NCBI\_TaxID=10407;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=14118, AND 11141;  
RX PubMed=10950984;  
RA Hannoun C., Norder H., Lindh M.;  
RT "An aberrant genotype revealed in recombinant hepatitis B virus strains from Vietnam.";

RT

RL J. Gen. Virol. 81:2267-2272(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=14118, AND 11141;  
RA Hannoun C., Norder H., Lindh M.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF241408; AAG17579.1; -  
DR EMBL; AF241407; AAG17572.1; -  
DR InterPro: IPR000236; TransactX.  
DR Pfam; PF00739; X; 1.  
SQ SEQUENCE 154 AA; 16516 MW; A13D1D623218640D CRC64;

Query Match 0.8%; Score 8; DB 12; Length 154;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GDLPSSSS 55  
Db 35 GDLPSSSS 42

RESULT 8  
Q9E6T5 PRELIMINARY; PRT; 154 AA.  
ID Q9E6T5  
AC Q9E6T5;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE X protein.  
GN X.  
OS Hepatitis B virus.  
OC Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.  
OX NCBI\_TaxID=10407;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=6871;  
RX MEDLINE=20409088; PubMed=10950984;  
RA Hannoun C., Norder H., Lindh M.;  
RT "An aberrant genotype revealed in recombinant hepatitis B virus strains from Vietnam.";  
RL J. Gen. Virol. 81:2267-2272(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=6871;  
RA Hannoun C., Norder H., Lindh M.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF241409; AAG17586.1; -  
DR InterPro: IPR000236; TransactX.  
DR Pfam; PF00739; X; 1.  
SQ SEQUENCE 154 AA; 16563 MW; E6B37B623218641B CRC64;

Query Match 0.8%; Score 8; DB 12; Length 154;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GDLPSSSS 55  
Db 35 GDLPSSSS 42

RESULT 9  
Q9ABT4 PRELIMINARY; PRT; 239 AA.  
ID Q9ABT4  
AC Q9ABT4;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein CC0136.  
GN CC0136.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
OC Caulobacter.  
OX NCBI\_TaxID=155892;



```

[1]
RN  SEQUENCE FROM N.A.
RP  STRAIN=ATCC 19089 / CB15;
RX  MEDLINE=21173698; PubMed=11259647;
RA  Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA  Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA  Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA  DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA  Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA  Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA  Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT  "Complete genome sequence of Caulobacter crescentus.";
RL  Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR  EMBL: AE005688; AAK22123.1; -.
DR  TIGR: CC0136; -.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 239 AA; 24704 MW; 1531A33ADACF9E31 CRC64;

Query Match      0.8%; Score 8; DB 16; Length 239;
Best Local Similarity 100.0%; Pred. No. 19;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  18 AEAPRRGR 25
Db  176 AEAPRRGR 183

RESULT 10
Q926W3
ID  Q926W3          PRELIMINARY;      PRT;      268 AA.
AC  Q926W3;
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DT  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT  01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE  Hypothetical protein lin2927.
GN  LIN2927.
OS  Listeria innocua.
OC  Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC  Listeriaceae; Listeria.
OX  NCBI_TaxID=1642;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=CLIP 11262 / SEROVAR 6A;
RX  PubMed=11679669;
RA  Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA  Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA  Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA  Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA  Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA  Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA  Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA  Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA  Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA  Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA  Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT  "Comparative genomics of Listeria species.";
RL  Science 294:849-852(2001).
DR  EMBL: AL596174; CAC98152.1; -.
DR  ListList; LIN02927; -.
DR  InterPro; IPR000281; HTH_RpIR.
DR  Pfam; PF01418; HTH_6; 1.
DR  Pfam; PF01380; SIS; 1.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 268 AA; 30656 MW; 7923711B8A818BE7 CRC64;

Query Match      0.8%; Score 8; DB 16; Length 268;
Best Local Similarity 100.0%; Pred. No. 21;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  792 LKNINIKE 799
Db  32 LKNINIKE 39

RESULT 11
Q8Y3P0
ID  Q8Y3P0          PRELIMINARY;      PRT;      268 AA.
AC  Q8Y3P0;
DT  01-MAR-2002 (TrEMBLrel. 20, Created)
DT  01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT  01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE  Hypothetical protein lmo2795.
GN  LMO2795.
OS  Listeria monocytogenes.
OC  Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC  Listeriaceae; Listeria.
OX  NCBI_TaxID=1639;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=EGD-E / SEROVAR 1/2A;
RX  MEDLINE=21537279; PubMed=11679669;
RA  Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA  Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA  Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA  Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA  Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA  Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA  Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA  Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA  Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA  Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA  Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT  "Comparative genomics of Listeria species.";
RL  Science 294:849-852(2001).
DR  EMBL: AL591984; CAD01008.1; -.
DR  ListList; LMO02795; -.
DR  InterPro; IPR000281; HTH_RpIR.
DR  InterPro; IPR001347; SIS.
DR  Pfam; PF01418; HTH_6; 1.
DR  Pfam; PF01380; SIS; 1.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 268 AA; 30643 MW; 7C26214E8FD1DAA2 CRC64;

Query Match      0.8%; Score 8; DB 16; Length 268;
Best Local Similarity 100.0%; Pred. No. 21;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  792 LKNINIKE 799
Db  32 LKNINIKE 39

RESULT 12
Q9TZN3
ID  Q9TZN3          PRELIMINARY;      PRT;      281 AA.
AC  Q9TZN3;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT  01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE  C43H8.3 protein.
GN  C43H8.3.
OS  Caenorhabditis elegans.
OC  Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC  Rhabditidae; Peloderinae; Caenorhabditis.
OX  NCBI_TaxID=6239;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BRISTOL N2;
RX  MEDLINE=94150718; PubMed=7906398;
RA  Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA  Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA  Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA  Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA  Jones M., Kershaw J., Kirsten J., Laister N., Latrelle P.,
RA  Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
```

```

[1]
RN  SEQUENCE FROM N.A.
RP  STRAIN=ATCC 19089 / CB15;
RX  MEDLINE=21173698; PubMed=11259647;
RA  Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA  Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA  Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA  DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA  Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA  Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA  Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT  "Complete genome sequence of Caulobacter crescentus.";
RL  Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR  EMBL: AE005688; AAK22123.1; -.
DR  TIGR: CC0136; -.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 239 AA; 24704 MW; 1531A33ADACF9E31 CRC64;

Query Match      0.8%; Score 8; DB 16; Length 239;
Best Local Similarity 100.0%; Pred. No. 19;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  18 AEAPRRGR 25
Db  176 AEAPRRGR 183

RESULT 10
Q926W3
ID  Q926W3          PRELIMINARY;      PRT;      268 AA.
AC  Q926W3;
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DT  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT  01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE  Hypothetical protein lin2927.
GN  LIN2927.
OS  Listeria innocua.
OC  Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC  Listeriaceae; Listeria.
OX  NCBI_TaxID=1642;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=CLIP 11262 / SEROVAR 6A;
RX  PubMed=11679669;
RA  Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA  Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA  Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA  Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA  Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA  Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA  Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA  Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA  Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA  Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA  Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT  "Comparative genomics of Listeria species.";
RL  Science 294:849-852(2001).
DR  EMBL: AL596174; CAC98152.1; -.
DR  ListList; LIN02927; -.
DR  InterPro; IPR000281; HTH_RpIR.
DR  Pfam; PF01418; HTH_6; 1.
DR  Pfam; PF01380; SIS; 1.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 268 AA; 30656 MW; 7923711B8A818BE7 CRC64;

Query Match      0.8%; Score 8; DB 16; Length 268;
Best Local Similarity 100.0%; Pred. No. 21;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  792 LKNINIKE 799
Db  32 LKNINIKE 39

RESULT 11
Q8Y3P0
ID  Q8Y3P0          PRELIMINARY;      PRT;      268 AA.
AC  Q8Y3P0;
DT  01-MAR-2002 (TrEMBLrel. 20, Created)
DT  01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT  01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE  Hypothetical protein lmo2795.
GN  LMO2795.
OS  Listeria monocytogenes.
OC  Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC  Listeriaceae; Listeria.
OX  NCBI_TaxID=1639;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=EGD-E / SEROVAR 1/2A;
RX  MEDLINE=21537279; PubMed=11679669;
RA  Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA  Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA  Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA  Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA  Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA  Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA  Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA  Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA  Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA  Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA  Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT  "Comparative genomics of Listeria species.";
RL  Science 294:849-852(2001).
DR  EMBL: AL591984; CAD01008.1; -.
DR  ListList; LMO02795; -.
DR  InterPro; IPR000281; HTH_RpIR.
DR  InterPro; IPR001347; SIS.
DR  Pfam; PF01418; HTH_6; 1.
DR  Pfam; PF01380; SIS; 1.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 268 AA; 30643 MW; 7C26214E8FD1DAA2 CRC64;

Query Match      0.8%; Score 8; DB 16; Length 268;
Best Local Similarity 100.0%; Pred. No. 21;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  792 LKNINIKE 799
Db  32 LKNINIKE 39

RESULT 12
Q9TZN3
ID  Q9TZN3          PRELIMINARY;      PRT;      281 AA.
AC  Q9TZN3;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT  01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE  C43H8.3 protein.
GN  C43H8.3.
OS  Caenorhabditis elegans.
OC  Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC  Rhabditidae; Peloderinae; Caenorhabditis.
OX  NCBI_TaxID=6239;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BRISTOL N2;
RX  MEDLINE=94150718; PubMed=7906398;
RA  Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA  Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA  Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA  Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA  Jones M., Kershaw J., Kirsten J., Laister N., Latrelle P.,
RA  Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
```

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
RA Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Tin-Wollam A.M.;  
RT "The sequence of C. elegans cosmid C43H8.";  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RL Submitted (OCT-1998) to the EMBL/GenBank/DBBJ databases.  
DR EMBL; AF098499; AAC67398.1; -.  
DR InterPro; IPR001791; Laminin\_G.  
DR Pfam; PF00054; Laminin\_G; 1.  
DR SMART; SM00282; LamG; 1.  
SQ SEQUENCE 281 AA; 31708 MW; 5253CC040277343C CRC64;

Query Match 0.8%; Score 8; DB 5; Length 281;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 NITDFTVK 738  
|||||  
Db 111 NITDFTVK 118

RESULT 13  
Q9PR14  
ID Q9PR14 PRELIMINARY; PRT; 286 AA.  
AC Q9PR14;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE ATP synthase gamma chain.  
GN. ATPG OR UUI30.  
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Ureaplasma.  
OX NCBI\_TaxID=134821;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SEROVAR 3;  
RX MEDLINE=20500219; PubMed=11048724;  
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,  
RA Cassell G.H.;  
RT "The complete sequence of the mucosal pathogen Ureaplasma  
urealyticum.";  
RL Nature 407:757-762(2000).  
DR EMBL; AE002114; AAF30536.1; -.  
DR InterPro; IPR000131; ATPase\_gamma.  
DR Pfam; PF00231; ATP-synt; 1.  
DR PRINTS; PR00126; ATPASEGAMMA.  
DR TIGRFAMs; TIGR01146; ATPsyn\_Flgamma; 1.  
KW Complete proteome.  
SQ SEQUENCE 286 AA; 32295 MW; C9AE278976AAE4F5 CRC64;

Query Match 0.8%; Score 8; DB 16; Length 286;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 EIVAGSDD 746  
|||||  
Db 279 EIVAGSDD 286

RESULT 14

Q8V9II  
ID Q8V9II PRELIMINARY; PRT; 293 AA.  
AC Q8V9II;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Movement protein.  
GN MP.  
OS Squash yellow mottle virus.  
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.  
OX NCBI\_TaxID=91372;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ramirez P., Karkashian J., Zuniga C., Maxwell D.;  
RT "Association of Squash Yellow Mottle Virus with cucurbits and papaya  
RT in Costa Rica.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBBJ databases.  
DR EMBL; AF440790; AAL33677.1; -.  
DR InterPro; IPR000211; Gemini\_BL.  
DR Pfam; PF00845; Gemini\_BL1; 1.  
SQ SEQUENCE 293 AA; 33078 MW; 8F99918F4FFE54AC CRC64;

Query Match 0.8%; Score 8; DB 12; Length 293;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 NYIESNRD 184  
|||||  
Db 13 NYIESNRD 20

RESULT 15  
Q50198  
ID Q50198 PRELIMINARY; PRT; 379 AA.  
AC Q50198;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE L222-ORF9.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97124199; PubMed=8969512;  
RA Fsihi H., De Rossi E., Salazar L., Cantoni R., Labo M., Riccardi G.,  
RA Takiff H.E., Eiglmeier K., Bergh S., Cole S.T.;  
RT "Gene arrangement and organization in a approximately 76 kb fragment  
RT encompassing the oric region of the chromosome of Mycobacterium  
RT leprae.";  
RL Microbiology 142:0-0(0).  
DR EMBL; L39923; AAB53129.1; -.  
DR InterPro; IPR004268; MVIN\_like.  
DR Pfam; PF03023; MVIN; 1.  
SQ SEQUENCE 379 AA; 40010 MW; FD23F9F20B4A74F5 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 379;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 TAVLLVAL 943  
|||||  
Db 243 TAVLLVAL 250

Search completed: May 12, 2003, 13:40:52  
Job time : 35.4548 secs







```
Qy 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
|||||
Db 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
|||||
Qy 121 SKCGEGTYSLGSGIKFDEWDELPAFESNIATFMDTVVGPSDSRDPGCGNNSSWIIPRGNYIE 180
|||||
Db 121 SKCGEGTYSLGSGIKFDEWDELPAFESNIATFMDTVVGPSDSRDPGCGNNSSWIIPRGNYIE 180
|||||
Qy 181 SNRDDCTVSLIYAVHLKKSgyVFFeYQYVDNNIFFEFFIQNDQCQEMDTTTDKWVKLTDN 240
|||||
Db 181 SNRDDCTVSLIYAVHLKKSgyVFFeYQYVDNNIFFEFFIQNDQCQEMDTTTDKWVKLTDN 240
|||||
Qy 241 GEWGSVMLKSGTNILYWRttGILMGSKAVKpVLvKNITIEGVAYTSECfPCKPGTFsN 300
|||||
Db 241 GEWGSVMLKSGTNILYWRttGILMGSKAVKpVLvKNITIEGVAYTSECfPCKPGTFsN 300
|||||
Qy 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYfQIHtPCDEEG 360
|||||
Db 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYfQIHtPCDEEG 360
|||||
Qy 361 KtQIMYKwIEPKICREDLTDAIRLPSGEEKKDCPPCNPNGFYNNSSSCHPCPPGTFSDGT 420
|||||
Db 361 KtQIMYKwIEPKICREDLTDAIRLPSGEEKKDCPPCNPNGFYNNSSSCHPCPPGTFSDGT 420
|||||
Qy 421 KECRPCPAGTEPALGFeyKwNvLPGNMKtSCFNvGNSKCDGMNGWEVAGDHIQSGAGGS 480
|||||
Db 421 KECRPCPAGTEPALGFeyKwNvLPGNMKtSCFNvGNSKCDGMNGWEVAGDHIQSGAGGS 480
|||||
Qy 481 DNDYLILNLHIPGfKpPtSMTGATGSELGRITfVfETLCSADCVLYfMvDINRKStNVVE 540
|||||
Db 481 DNDYLILNLHIPGfKpPtSMTGATGSELGRITfVfETLCSADCVLYfMvDINRKStNVVE 540
|||||
Qy 541 SWGGTKEKQAYTHIIFKNATfTfTWAFQRTNQGDNRrFINdMVkiYSITATNAVDGvAS 600
|||||
Db 541 SWGGTKEKQAYTHIIFKNATfTfTWAFQRTNQGDNRrFINdMVkiYSITATNAVDGvAS 600
|||||
Qy 601 SCRACALGSEQSGSSCVPCPPGHYIEKtEQKecPPDTYLSIHQVYGKEACIPCGPGSK 660
|||||
Db 601 SCRACALGSEQSGSSCVPCPPGHYIEKtEQKecPPDTYLSIHQVYGKEACIPCGPGSK 660
|||||
Qy 661 NNQDHSVCSDCFFYHEKENQILHvDfSNLSSVGLMNGPStfSKGTkYfHfFNISLCGH 720
|||||
Db 661 NNQDHSVCSDCFFYHEKENQILHvDfSNLSSVGLMNGPStfSKGTkYfHfFNISLCGH 720
|||||
Qy 721 EGKKMALCTNNITDfTVKEIvAGSDDYtNLvGAFvCQStIIPSESKGFRAALSSQSIILA 780
|||||
Db 721 EGKKMALCTNNITDfTVKEIvAGSDDYtNLvGAFvCQStIIPSESKGFRAALSSQSIILA 780
|||||
Qy 781 DTFIGTVtETTLKNINIKEDMfPVPtSQIPDVHfFYKStATTSCINGrStAVKMRCNPT 840
|||||
Db 781 DTFIGTVtETTLKNINIKEDMfPVPtSQIPDVHfFYKStATTSCINGrStAVKMRCNPT 840
|||||
Qy 841 KSGAGVISVPSKCPAGTCdGCTfYFLWESAeACPLCTEHDFHEIeGACKRGfQETLYvWN 900
|||||
Db 841 KSGAGVISVPSKCPAGTCdGCTfYFLWESAeACPLCTEHDFHEIeGACKRGfQETLYvWN 900
|||||
Qy 901 EPKwCIKGIStLPEKKLatCETvDfWLKvGAGvGAfTAvLLvALtCYfWKKNOKLEYKYSK 960
|||||
Db 901 EPKwCIKGIStLPEKKLatCETvDfWLKvGAGvGAfTAvLLvALtCYfWKKNOKLEYKYSK 960
|||||
Qy 961 LVMTTNSKECELPAAADSCAIeGEDNEEEVvYSNKQSLlGLKSLATKEKEDHFESVQLK 1020
|||||
Db 961 LVMTTNSKECELPAAADSCAIeGEDNEEEVvYSNKQSLlGLKSLATKEKEDHFESVQLK 1020
|||||
Qy 1021 TSRSpNI 1027
|||||
Db 1021 TSRSpNI 1027
|||||
```

RESULT 2  
US-10-140-164-2  
; Sequence 2, Application US/10140164  
; Publication No. US20030072736A1

```
; GENERAL INFORMATION:  
; APPLICANT: Baker et al.  
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16  
; FILE REFERENCE: PF514C1  
; CURRENT APPLICATION NUMBER: US/10/140,164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 963  
; TYPE: prt  
; ORGANISM: Homo sapiens  
; US-10-140-164-2
```

```
Query Match 92.8%; Score 953; DB 9; Length 963;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 953; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 MLFRARGPVRGRCWRPAEAPRRGRSPWPSPAWICCWALAGCAAAGDLPSSSSRPLPP 60
|||||
Db 1 MLFRARGPVRGRCWRPAEAPRRGRSPWPSPAWICCWALAGCAAAGDLPSSSSRPLPP 60
|||||
Qy 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
|||||
Db 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
|||||
Qy 121 SKCGEGTYSLGSGIKFDEWDELPAFESNIATFMDTVVGPSDSRDPGCGNNSSWIIPRGNYIE 180
|||||
Db 121 SKCGEGTYSLGSGIKFDEWDELPAFESNIATFMDTVVGPSDSRDPGCGNNSSWIIPRGNYIE 180
|||||
Qy 181 SNRDDCTVSLIYAVHLKKSgyVFFeYQYVDNNIFFEFFIQNDQCQEMDTTTDKWVKLTDN 240
|||||
Db 181 SNRDDCTVSLIYAVHLKKSgyVFFeYQYVDNNIFFEFFIQNDQCQEMDTTTDKWVKLTDN 240
|||||
Qy 241 GEWGSVMLKSGTNILYWRttGILMGSKAVKpVLvKNITIEGVAYTSECfPCKPGTFsN 300
|||||
Db 241 GEWGSVMLKSGTNILYWRttGILMGSKAVKpVLvKNITIEGVAYTSECfPCKPGTFsN 300
|||||
Qy 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYfQIHtPCDEEG 360
|||||
Db 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYfQIHtPCDEEG 360
|||||
Qy 361 KtQIMYKwIEPKICREDLTDAIRLPSGEEKKDCPPCNPNGFYNNSSSCHPCPPGTFSDGT 420
|||||
Db 361 KtQIMYKwIEPKICREDLTDAIRLPSGEEKKDCPPCNPNGFYNNSSSCHPCPPGTFSDGT 420
|||||
Qy 421 KECRPCPAGTEPALGFeyKwNvLPGNMKtSCFNvGNSKCDGMNGWEVAGDHIQSGAGGS 480
|||||
Db 421 KECRPCPAGTEPALGFeyKwNvLPGNMKtSCFNvGNSKCDGMNGWEVAGDHIQSGAGGS 480
|||||
Qy 481 DNDYLILNLHIPGfKpPtSMTGATGSELGRITfVfETLCSADCVLYfMvDINRKStNVVE 540
|||||
Db 481 DNDYLILNLHIPGfKpPtSMTGATGSELGRITfVfETLCSADCVLYfMvDINRKStNVVE 540
|||||
Qy 541 SWGGTKEKQAYTHIIFKNATfTfTWAFQRTNQGDNRrFINdMVkiYSITATNAVDGvAS 600
|||||
Db 541 SWGGTKEKQAYTHIIFKNATfTfTWAFQRTNQGDNRrFINdMVkiYSITATNAVDGvAS 600
|||||
```

Qy 601 SCRACALGSEQSGSCVPCPPGHYIEKETNQCKECPDPTYLSIHQVYGKEACIPCGPGSK 660  
Db 601 SCRACALGSEQSGSCVPCPPGHYIEKETNQCKECPDPTYLSIHQVYGKEACIPCGPGSK 660  
Qy 661 NNQDHSVYSDCFFYHEKENQILHYDFSNLSSVGLMNGPSEFTSGTKYFHFNISLCGH 720  
Db 661 NNQDHSVYSDCFFYHEKENQILHYDFSNLSSVGLMNGPSEFTSGTKYFHFNISLCGH 720  
Qy 721 EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFCVQCSTIIPSESKGFRAALSSQSIILA 780  
Db 721 EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFCVQCSTIIPSESKGFRAALSSQSIILA 780  
Qy 781 DTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSSTATTSCTINGRSTAVKMRNCPT 840  
Db 781 DTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSSTATTSCTINGRSTAVKMRNCPT 840  
Qy 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAECPLCTEHDHFHEIEGACKRGFQETLYVWN 900  
Db 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAECPLCTEHDHFHEIEGACKRGFQETLYVWN 900  
Qy 901 EPKWCIGISLPEKKLATCETVDFWLKVGAGVGAGTAVALLLVALTCYFWKKNQK 953  
Db 901 EPKWCIGISLPEKKLATCETVDFWLKVGAGVGAGTAVALLLVALTCYFWKKNQK 953  
RESULT 3  
US-10-002-050-20  
; Sequence 20, Application US/10002050  
; Publication No. US20030032095A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: No: US20030032095A1el Nucleic Acid Sequences Encoding Human Sema  
; FILE REFERENCE: 15966-554 Cura-54 CON-S14  
; CURRENT APPLICATION NUMBER: US/10/002,050  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 464  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-002-050-20

Query Match 26.4%; Score 271; DB 9; Length 464;  
Best Local Similarity 99.6%; Pred. No. 1.2e-260;  
Matches 451; Conservative 0; Mismatches 0; Indels 2; Gaps 1;  
Qy 115 MKNQVCSKCGEGTYSLGSGIKFDEWDELPAFNSNIATFMDTVVGPDSRDPDGCNNSSWIP 174  
Db 1 MKNQVCSKCGEGTYSLGSGIKFDEWDELPAFNSNIATFMDTVVGPDSRDPDGCNNSSWIP 60  
Qy 175 RGNYESNRDDCTVSLIYAVHLKKSgyvffeyqyvDNNIFFEFFIIONDQCQEMDTTTDKW 234  
Db 61 RGNYESNRDDCTVSLIYAVHLKKSgyvffeyqyvDNNIFFEFFIIONDQCQEMDTTTDKW 120  
Qy 235 VKLTDNGEGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCK 294  
Db 121 VKLTDNGEGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCK 180  
Qy 295 PGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTERPPCTTKDYFQI 352  
Db 61 RGNYESNRDDCTVSLIYAVHLKKSgyvffeyqyvDNNIFFEFFIIONDQCQEMDTTTDKW 120  
Qy 235 VKLTDNGEGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCK 294  
Db 121 VKLTDNGEGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCK 180  
Qy 295 PGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTERPPCTTKDYFQI 352  
Db 181 PGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTERPPCTTKDYFQI 240

Qy 353 HTPCDEEGKTQIMYKWIPEKICREDLTDAIRLPPSGEKKDCPPCNPNGFYNNGSSSCHPCP 412  
Db 241 HTPCDEEGKTQIMYKWIPEKICREDLTDAIRLPPSGEKKDCPPCNPNGFYNNGSSSCHPCP 300  
Qy 413 PGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 472  
Db 301 PGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 360  
Qy 473 IQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITVFETLCSADCVLYFMVDIN 532  
Db 361 IQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITVFETLCSADCVLYFMVDIN 420  
Qy 533 RKSTNVVESWGTKEKQAYTHIIFKNATFTFTW 565  
Db 421 RKSTNVVESWGTKEKQAYTHIIFKNATFTFTW 453  
RESULT 4  
US-10-002-304-20  
; Sequence 20, Application US/10002304  
; Publication No. US20030036185A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby  
; FILE REFERENCE: 15966-554 Cura-54 CON-S8  
; CURRENT APPLICATION NUMBER: US/10/002,304  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 464  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-002-304-20

Query Match 26.4%; Score 271; DB 9; Length 464;  
Best Local Similarity 99.6%; Pred. No. 1.2e-260;  
Matches 451; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

Qy 115 MKNQVCSKCGEGTYSLGSGIKFDEWDELPAFNSNIATFMDTVVGPDSRDPDGCNNSSWIP 174  
Db 1 MKNQVCSKCGEGTYSLGSGIKFDEWDELPAFNSNIATFMDTVVGPDSRDPDGCNNSSWIP 60  
Qy 175 RGNYESNRDDCTVSLIYAVHLKKSgyvffeyqyvDNNIFFEFFIIONDQCQEMDTTTDKW 234  
Db 61 RGNYESNRDDCTVSLIYAVHLKKSgyvffeyqyvDNNIFFEFFIIONDQCQEMDTTTDKW 120  
Qy 235 VKLTDNGEGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCK 294  
Db 121 VKLTDNGEGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCK 180  
Qy 295 PGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTERPPCTTKDYFQI 352  
Db 181 PGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTERPPCTTKDYFQI 240  
Qy 353 HTPCDEEGKTQIMYKWIPEKICREDLTDAIRLPPSGEKKDCPPCNPNGFYNNGSSSCHPCP 412  
Db 241 HTPCDEEGKTQIMYKWIPEKICREDLTDAIRLPPSGEKKDCPPCNPNGFYNNGSSSCHPCP 300  
Qy 413 PGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 472  
Db 301 PGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 360  
Qy 473 IQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITVFETLCSADCVLYFMVDIN 532

Db 361 IQSGAGGSDNDYLILNLHIPGKPPPTSMGTGATGSELGRITFVFETLCSADCVLYFMVDIN 420  
QY 533 RKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 565  
Db 421 RKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 453  
RESULT 5  
US-10-003-152-20  
; Sequence 20, Application US/10003152  
; Patent No. US20020151494A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: No. US20020151494A1el Amino Acid Sequences for Human Semaphorin-1  
; FILE REFERENCE: 15966-554 Cura-54 CON-S12  
; CURRENT APPLICATION NUMBER: US/10/003,152  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 464  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-003-152-20

US-10-002-050-10  
; Sequence 10, Application US/10002050  
; Publication No. US20030032095A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: No. US20030032095A1el Nucleic Acid Sequences Encoding Human Se  
; FILE REFERENCE: 15966-554 Cura-54 CON-S14  
; CURRENT APPLICATION NUMBER: US/10/002,050  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 411  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-002-050-10  
Query Match 22.6%; Score 232; DB 9; Length 411;  
Best Local Similarity 100.0%; Pred. No. 6.4e-222;  
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 334 GSSECTERPPCTTKDYFOIHTPCDEEGTKQIMYKWIIEPKICREDLTDAIRLPPSGEKKDC 393  
Db 169 GSSECTERPPCTTKDYFOIHTPCDEEGTKQIMYKWIIEPKICREDLTDAIRLPPSGEKKDC 228  
QY 394 PPCNPGFYNNSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFEYKWNVLPGNMKTSCF 453  
Db 229 PPCNPGFYNNSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFEYKWNVLPGNMKTSCF 288  
QY 454 NVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGKPPPTSMGTGATGSELGRITF 513  
Db 289 NVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGKPPPTSMGTGATGSELGRITF 348  
QY 514 VFETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 565  
Db 349 VFETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 400  
RESULT 7  
US-10-002-304-10  
; Sequence 10, Application US/10002304  
; Publication No. US20030036185A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby  
; FILE REFERENCE: 15966-554 Cura-54 CON-S8  
; CURRENT APPLICATION NUMBER: US/10/002,304  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 411  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-002-304-10

QY 115 MKNQVCSKCGEGTSLYAVHLKKGSGYVFEYQYVDNFIFFEQNDQCEMDTTTDKW 174  
Db 1 MKNQVCSKCGEGTSLYAVHLKKGSGYVFEYQYVDNFIFFEQNDQCEMDTTTDKW 60  
QY 175 RGNYESNRDCTVSLIYAVHLKKGSGYVFEYQYVDNFIFFEQNDQCEMDTTTDKW 234  
Db 61 RGNYESNRDCTVSLIYAVHLKKGSGYVFEYQYVDNFIFFEQNDQCEMDTTTDKW 120  
QY 235 VKLTDNGEWSHVMKSGTNILYRRTGILMGSKAVKPVLVKNITIEGVAYTSECFPC 294  
Db 121 VKLTDNGEWSHVMKSGTNILYRRTGILMGSKAVKPVLVKNITIEGVAYTSECFPC 180  
QY 295 PGTFESNKPQSFNCQVCPRNTYSEKGAKECIRCKDDSDQSFS--GSSECTERPPCTTKDYFOI 352  
Db 181 PGTFESNKPQSFNCQVCPRNTYSEKGAKECIRCKDDSDQSFSSEGSSECTERPPCTTKDYFOI 240  
QY 353 HTPCDEEGTKQIMYKWIIEPKICREDLTDAIRLPPSGEKKDCPPCNPFGYNNSSSCHPCP 412  
Db 241 HTPCDEEGTKQIMYKWIIEPKICREDLTDAIRLPPSGEKKDCPPCNPFGYNNSSSCHPCP 300  
QY 413 PGTFSDGTKECRPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 472  
Db 301 PGTFSDGTKECRPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 360  
QY 473 IQSGAGGSDNDYLILNLHIPGKPPPTSMGTGATGSELGRITFVFETLCSADCVLYFMVDIN 532  
Db 361 IQSGAGGSDNDYLILNLHIPGKPPPTSMGTGATGSELGRITFVFETLCSADCVLYFMVDIN 420  
QY 533 RKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 565  
Db 421 RKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 453



```
Query Match      22.6%; Score 232; DB 9; Length 411;
Best Local Similarity 100.0%; Pred. No. 6.4e-222;
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 GSSECTERPPCTTKDYFOIHTPCDEEGKTQIMYKWIETKICREDLTDAILRPPSGEKKDC 393
Db 169 GSSECTERPPCTTKDYFOIHTPCDEEGKTQIMYKWIETKICREDLTDAILRPPSGEKKDC 228

QY 394 PPCNPGFYNGSSSCHPCPPGTFSDGTKECRPCPCAGTEPALGFEYKWNVLPGNMKTSCF 453
Db 229 PPCNPGFYNGSSSCHPCPPGTFSDGTKECRPCPCAGTEPALGFEYKWNVLPGNMKTSCF 288

QY 454 NVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITF 513
Db 289 NVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITF 348

QY 514 VFETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 565
Db 349 VFETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 400

RESULT 8
US-10-003-152-10
; Sequence 10, Application US/10003152
; Patent No. US20020151494A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meijia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20020151494A1el Amino Acid Sequences for Human Semaphorin-I
; FILE REFERENCE: 15966-554 Cura-54 CON-S12
; CURRENT APPLICATION NUMBER: US/10/003,152
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-003-152-10

Query Match      22.6%; Score 232; DB 12; Length 411;
Best Local Similarity 100.0%; Pred. No. 6.4e-222;
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 GSSECTERPPCTTKDYFOIHTPCDEEGKTQIMYKWIETKICREDLTDAILRPPSGEKKDC 393
Db 169 GSSECTERPPCTTKDYFOIHTPCDEEGKTQIMYKWIETKICREDLTDAILRPPSGEKKDC 228

QY 394 PPCNPGFYNGSSSCHPCPPGTFSDGTKECRPCPCAGTEPALGFEYKWNVLPGNMKTSCF 453
Db 229 PPCNPGFYNGSSSCHPCPPGTFSDGTKECRPCPCAGTEPALGFEYKWNVLPGNMKTSCF 288

QY 454 NVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITF 513
Db 289 NVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITF 348

QY 514 VFETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 565
Db 349 VFETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 400

RESULT 9
US-09-864-761-39769
; Sequence 39769, Application US/09864761
```

```
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL, FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39769
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002081.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EST_HUMAN HIT: AW954806.1, EVALUE 5.00e-45
; OTHER INFORMATION: SWISSPROT HIT: P01267, EVALUE 3.00e-03
US-09-864-761-39769

Query Match      7.9%; Score 81; DB 10; Length 81;
Best Local Similarity 100.0%; Pred. No. 1.9e-72;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGGHYIEKETNQKCECPP 637
Db 1 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGGHYIEKETNQKCECPP 60
```



QY 638 DTYLSIHQVYGKEACIPCGPG 658  
Db 61 DTYLSIHQVYGKEACIPCGPG 81

RESULT 10  
US-10-140-164-36  
; Sequence 36, Application US/10140164  
; Publication No. US20030072736A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker et al.  
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16  
; FILE REFERENCE: PF514C1  
; CURRENT APPLICATION NUMBER: US/10/140,164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 36  
; LENGTH: 78  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-140-164-36

Query Match 7.6%; Score 78; DB 9; Length 78;  
Best Local Similarity 100.0%; Pred. No. 1.7e-69;  
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 950 KNQKLEYKYSKLVMTTNSKECELPAADSCAIMEGEDNEEEVVYSNKSLLGKLSLATKE 1009  
Db 1 KNQKLEYKYSKLVMTTNSKECELPAADSCAIMEGEDNEEEVVYSNKSLLGKLSLATKE 60

QY 1010 KEDHFESVQLKTSRSPNI 1027  
Db 61 KEDHFESVQLKTSRSPNI 78

RESULT 11  
US-10-140-164-65  
; Sequence 65, Application US/10140164  
; Publication No. US20030072736A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker et al.  
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16  
; FILE REFERENCE: PF514C1  
; CURRENT APPLICATION NUMBER: US/10/140,164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 36  
; LENGTH: 78  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-140-164-36

; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 65  
; LENGTH: 78  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-140-164-65

Query Match 7.6%; Score 78; DB 9; Length 78;  
Best Local Similarity 100.0%; Pred. No. 1.7e-69;  
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 950 KNQKLEYKYSKLVMTTNSKECELPAADSCAIMEGEDNEEEVVYSNKSLLGKLSLATKE 1009  
Db 1 KNQKLEYKYSKLVMTTNSKECELPAADSCAIMEGEDNEEEVVYSNKSLLGKLSLATKE 60

QY 1010 KEDHFESVQLKTSRSPNI 1027  
Db 61 KEDHFESVQLKTSRSPNI 78

RESULT 12  
US-10-140-164-32  
; Sequence 32, Application US/10140164  
; Publication No. US20030072736A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker et al.  
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16  
; FILE REFERENCE: PF514C1  
; CURRENT APPLICATION NUMBER: US/10/140,164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 32  
; LENGTH: 74  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-140-164-32

Query Match 7.2%; Score 74; DB 9; Length 74;  
Best Local Similarity 100.0%; Pred. No. 1.6e-65;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 CTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIPEKICREDLTDAIRLPPSGEKKDCPPCN 397  
Db 1 CTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIPEKICREDLTDAIRLPPSGEKKDCPPCN 60

QY 398 PGFYNNGSSSCHPC 411  
Db 61 PGFYNNGSSSCHPC 74

RESULT 13  
US-10-140-164-61  
; Sequence 61, Application US/10140164  
; Publication No. US20030072736A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker et al.  
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16  
; FILE REFERENCE: PF514C1  
; CURRENT APPLICATION NUMBER: US/10/140,164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 61  
; LENGTH: 74  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-140-164-61

Query Match 7.2%; Score 74; DB 9; Length 74;  
Best Local Similarity 100.0%; Pred. No. 1.6e-65;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 338 CTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIETKICREDLTDAIRLPPSGEKKDCPPCN 397  
Db 1 CTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIETKICREDLTDAIRLPPSGEKKDCPPCN 60

Qy 398 PGFYNNGSSSCHPC 411  
Db 61 PGFYNNGSSSCHPC 74

RESULT 14  
US-09-864-761-47095  
; Sequence 47095, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aeomica-x-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 47095  
; LENGTH: 64  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC002081.1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.55  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.52  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.49  
; OTHER INFORMATION: SWISSPROT HIT: Q00019, EVALUE 8.90e-01  
US-09-864-761-47095

Query Match 6.2%; Score 64; DB 10; Length 64;  
Best Local Similarity 100.0%; Pred. No. 1.2e-55;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 FSCASGEYLEMKNQVCSKCGEGTYSLGSGIKFDEWDELPAFGFSNIATFMDTVVGPDSR 164  
Db 1 FSCASGEYLEMKNQVCSKCGEGTYSLGSGIKFDEWDELPAFGFSNIATFMDTVVGPDSR 60

Qy 165 DGCN 168  
Db 61 DGCN 64

RESULT 15  
US-09-864-761-39057  
; Sequence 39057, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aeomica-x-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6

```
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39057
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002081.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
; OTHER INFORMATION: EST_HUMAN HIT: AW966212.1, EVALUE 2.00e-10
; OTHER INFORMATION: SWISSPROT HIT: P21849, EVALUE 5.00e-03
US-09-864-761-39057

Query Match          5.8%; Score 60; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.1e-51;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 TOIMYKWIEPKICREDLTDAIRLPPSGEKKDCPPCNPFGYNNNGSSSCHPCPPGTFSDGTK 421
   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 TOIMYKWIEPKICREDLTDAIRLPPSGEKKDCPPCNPFGYNNNGSSSCHPCPPGTFSDGTK 60
```

Search completed: May 12, 2003, 13:49:26  
Job time : 23.1271 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:23:08 ; Search time 36.1256 Seconds  
(without alignments)  
3788.125 Million cell updates/sec

Title: US-10-073-333A-4  
Perfect score: 1027  
Sequence: 1 MLFRARGPVRRGWRPAEA.....KEKEDHFESVQLKTSRSPNI 1027

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1027	100.0	1027	22 AAB70256	TR16-long receptor
2	953	92.8	963	22 AAB70255	TR16-short recepto
3	271	26.4	464	22 AAB48377	Human SEC10 protei
4	232	22.6	411	22 AAB48372	Human SEC5 protein
5	81	7.9	81	22 ABB39918	Peptide #7424 enco
6	81	7.9	81	22 ABB24471	Protein #6470 enco
7	81	7.9	81	22 AAM60663	Human brain expres
8	81	7.9	81	22 AAM73335	Human bone marrow
9	81	7.9	81	22 AAM33535	Peptide #7572 enco
10	81	7.9	81	23 ABG43186	Human peptide enco

11	78	7.6	78	22 AAB70285	Peptide #29. Unid
12	74	7.2	74	22 AAB70281	Peptide #25. Unid
13	64	6.2	64	22 AAM72925	Human bone marrow
14	64	6.2	64	23 ABG42760	Human peptide enco
15	60	5.8	60	22 ABB38686	Peptide #6192 enco
16	60	5.8	60	22 ABB23759	Protein #5758 enco
17	60	5.8	60	22 AAM59318	Human brain expres
18	60	5.8	60	22 AAM71867	Human bone marrow
19	60	5.8	60	22 AAM32149	Peptide #6186 enco
20	60	5.8	60	23 ABG41680	Human peptide enco
21	50	4.9	50	22 ABB39681	Peptide #7187 enco
22	50	4.9	50	22 ABB24346	Protein #6345 enco
23	50	4.9	50	22 AAM60397	Human brain expres
24	50	4.9	50	22 AAM73033	Human bone marrow
25	50	4.9	50	22 AAM19811	Peptide #6245 enco
26	50	4.9	50	22 AAM33257	Peptide #7294 enco
27	50	4.9	50	23 ABG42877	Human peptide enco
28	37	3.6	71	22 AAU21345	Human novel foetal
29	21	2.0	105	21 AAB26180	Human CASB619 prot
30	21	2.0	495	20 AAY59972	Human endometrium
31	21	2.0	750	22 AAB35328	Human TR13 recepto
32	21	2.0	1001	22 AAB35333	Human TR13 recepto
33	21	2.0	1013	21 AAB26179	Human CASB619 prot
34	21	2.0	1013	22 AAU12190	Human PRO4985 poly
35	21	2.0	1013	22 AAB83845	Amino acid sequenc
36	15	1.5	372	22 AAB85768	Human seven-transm
37	15	1.5	383	22 AAB83853	Amino acid sequenc
38	15	1.5	870	22 AAB83851	Amino acid sequenc
39	15	1.5	911	22 AAB83850	Amino acid sequenc
40	11	1.1	52	22 AAB83848	Peptide fragment o
41	11	1.1	209	22 AAB83852	Amino acid sequenc
42	9	0.9	9	21 AAB26190	Human CASB619 prot
43	9	0.9	212	22 ABG04843	Novel human diagno
44	9	0.9	708	22 AAG91270	C glutamicum prote
45	9	0.9	1487	19 AAW76719	Thermococcus 9N2 b

ALIGNMENTS

RESULT 1  
AAB70256  
ID AAB70256 standard; protein; 1027 AA.  
XX  
AC AAB70256;  
XX  
DT 10-MAY-2001 (first entry)  
XX  
DE TR16-long receptor protein.  
XX  
KW TR16 recptor; tumour necrosis factor receptor superfamily;  
KW apoptosis; inflammatory; cancer; immune; neurodegenerative.  
XX  
OS Unidentified.  
XX  
PN WO200112671-A1.  
XX  
PD 22-FEB-2001.  
XX  
PF 10-AUG-2000; 2000WO-US21885.  
XX  
PR 12-AUG-1999; 99US-0148348.  
PR 13-AUG-1999; 99US-0148683.  
PR 13-AUG-1999; 99US-0148870.  
PR 16-AUG-1999; 99US-0148758.  
PR 17-AUG-1999; 99US-0149181.  
PR 18-AUG-1999; 99US-0149453.  
PR 19-AUG-1999; 99US-0149498.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Young PE, Baker KP;  
XX

DR	WPI; 2001-138754/14.	
XX	New nucleic acid molecule encoding a TR16 tumor necrosis factor	
PT	receptor polypeptide, useful for the diagnosis and treatment of cancer,	
PT	autoimmune disorders and cardiovascular diseases -	
XX		
PS	Disclosure; Fig 4; 286pp; English.	
XX		
CC	The present invention relates to a TR16 receptor (tumour necrosis	
CC	factor receptor superfamily). The invention is useful treating	
CC	diseases and disorders associated with the inhibited or increased	
CC	apoptosis. In particular inflammatory diseases, cancers, immune and	
CC	neurodegenerative disorders may be treated.	
XX		
SQ	Sequence 1027 AA;	
Query Match 100.0%; Score 1027; DB 22; Length 1027;		
Best Local Similarity 100.0%; Pred. No. 0;		
Matches 1027; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 MLFRARGPVRGRGWGRPAEAPRRGRSPWPSPAWICCWALAGCAAAGDLPSSSSRPLPP 60	
Db	1 MLFRARGPVRGRGWGRPAEAPRRGRSPWPSPAWICCWALAGCAAAGDLPSSSSRPLPP 60	
QY	61 CQEKDYHFYEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120	
Db	61 CQEKDYHFYEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120	
QY	121 SKCGEGTYSLGSIGIKFDEWDELPAFNSNIATFMDTVVGPDSRDPGCGNNSWIIPRGNYIE 180	
Db	121 SKCGEGTYSLGSIGIKFDEWDELPAFNSNIATFMDTVVGPDSRDPGCGNNSWIIPRGNYIE 180	
QY	181 SNRDCTVSLIYAVHLKKSGYVFEFYQYVDNNIFFEYFIQNDQCQEMDTTDDKWKVCLTDN 240	
Db	181 SNRDCTVSLIYAVHLKKSGYVFEFYQYVDNNIFFEYFIQNDQCQEMDTTDDKWKVCLTDN 240	
QY	241 GEWGSVSVMLKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECFPCPKPGTFSN 300	
Db	241 GEWGSVSVMLKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECFPCPKPGTFSN 300	
QY	301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFOIHTPCDEEG 360	
Db	301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFOIHTPCDEEG 360	
QY	361 KTOIMYKWIETPKICREDLTDALRPPSGEKKDCRPPCNPGFYNNGSSSCHPCPPGTFSDGT 420	
Db	361 KTOIMYKWIETPKICREDLTDALRPPSGEKKDCRPPCNPGFYNNGSSSCHPCPPGTFSDGT 420	
QY	421 KECRPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480	
Db	421 KECRPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480	
QY	481 DNDYLILNLHIPGFKPPTSMGTGATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540	
Db	481 DNDYLILNLHIPGFKPPTSMGTGATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540	
QY	541 SWGGTKEKQAYTHIIFKNATFTTFAFQRTNQGDNRFRFINDMVKIYSITATNAVVGVAS 600	
Db	541 SWGGTKEKQAYTHIIFKNATFTTFAFQRTNQGDNRFRFINDMVKIYSITATNAVVGVAS 600	
QY	601 SCRACALGSEQSGSSCVPCPPGHVIEKETNQCKECPDPTVLSIHQVYGKEACIPCGPGSK 660	
Db	601 SCRACALGSEQSGSSCVPCPPGHVIEKETNQCKECPDPTVLSIHQVYGKEACIPCGPGSK 660	
QY	661 NNQDHSVYSDCFYHEKENQILHYDFSNLSSVGLMNGPSTSGTKYFHFENISLCGH 720	
Db	661 NNQDHSVYSDCFYHEKENQILHYDFSNLSSVGLMNGPSTSGTKYFHFENISLCGH 720	
QY	721 EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFCVQCSTIIPSESKGFRAALSSQSIILA 780	
Db	721 EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFCVQCSTIIPSESKGFRAALSSQSIILA 780	
QY	781 DTFIGTVVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSSTATTSINGRSTAVKMRCNPT 840	

Db	781 DTFIGTVVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSSTATTSINGRSTAVKMRCNPT 840	
QY	841 KSGAGVISVPSKCPAGTCDGCTFFFLWESAEACPLCTEHDFHEIEGACKRGFQETLYVWN 900	
Db	841 KSGAGVISVPSKCPAGTCDGCTFFFLWESAEACPLCTEHDFHEIEGACKRGFQETLYVWN 900	
QY	901 EPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAFVALLVALTCYFWKKNQKLEYKYSK 960	
Db	901 EPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAFVALLVALTCYFWKKNQKLEYKYSK 960	
QY	961 LVMTTNSKECELPAAADSCAIMEGEDNEEEVVSNNKQSLGLKLSLATKEKEDHFESVQLK 1020	
Db	961 LVMTTNSKECELPAAADSCAIMEGEDNEEEVVSNNKQSLGLKLSLATKEKEDHFESVQLK 1020	
QY	1021 TSRSNPNI 1027	
Db	1021 TSRSNPNI 1027	
RESULT 2		
AAB70255		
ID	AAB70255 standard; protein; 963 AA.	
XX		
AC	AAB70255;	
XX		
DT	10-MAY-2001 (first entry)	
XX		
DE	TR16-short receptor protein.	
XX		
KW	TR16 receptor; tumour necrosis factor receptor superfamily;	
KW	apoptosis; inflammatory; cancer; immune; neurodegenerative.	
XX		
OS	Unidentified.	
XX		
PN	WO200112671-A1.	
XX		
PD	22-FEB-2001.	
XX		
PF	10-AUG-2000; 2000WO-US21885.	
XX		
PR	12-AUG-1999; 99US-0148348.	
PR	13-AUG-1999; 99US-0148683.	
PR	13-AUG-1999; 99US-0148870.	
PR	16-AUG-1999; 99US-0148758.	
PR	17-AUG-1999; 99US-0149181.	
PR	18-AUG-1999; 99US-0149453.	
PR	19-AUG-1999; 99US-0149498.	
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Ruben SM, Young PE, Baker KP;	
XX		
DR	WPI; 2001-138754/14.	
XX		
PT	New nucleic acid molecule encoding a TR16 tumor necrosis factor	
PT	receptor polypeptide, useful for the diagnosis and treatment of cancer,	
PT	autoimmune disorders and cardiovascular diseases -	
XX		
PS	Claim 1; Fig 1; 286pp; English.	
XX		
CC	The present invention relates to a TR16 receptor (tumour necrosis	
CC	factor receptor superfamily). The invention is useful treating	
CC	diseases and disorders associated with the inhibited or increased	
CC	apoptosis. In particular inflammatory diseases, cancers, immune and	
CC	neurodegenerative disorders may be treated.	
XX		
SQ	Sequence 963 AA;	

Query Match 92.8%; Score 953; DB 22; Length 963;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 953; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFRARGPVGRGWGRPAEAPRRGRSPWPSPAWICCWALAGCQAAWAGDLPSSSSRPLPP 60  
Db 1 MLFRAGPVGRGWGRPAEAPRRGRSPWPSPAWICCWALAGCQAAWAGDLPSSSSRPLPP 60  
QY 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120  
Db 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120  
QY 121 SKCGEGTSLGSGIKFDEWDELPAFESNIATFMDTVVGPSDSRDPGCNNSSWIIPRGNYIE 180  
Db 121 SKCGEGTSLGSGIKFDEWDELPAFESNIATFMDTVVGPSDSRDPGCNNSSWIIPRGNYIE 180  
QY 181 SNRDDCTVSLIYAVHLKKSGYVFFEQYVDNNIFFEQQNDQCQEMDTTDKWKLTDN 240  
Db 181 SNRDDCTVSLIYAVHLKKSGYVFFEQYVDNNIFFEQQNDQCQEMDTTDKWKLTDN 240  
QY 241 GEWGSVMLKSGTNIYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFFCKPGTFSN 300  
Db 241 GEWGSVMLKSGTNIYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFFCKPGTFSN 300  
QY 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360  
Db 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360  
QY 361 KTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPNGFYNNSSSCHPCPPGTFSDGT 420  
Db 361 KTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPNGFYNNSSSCHPCPPGTFSDGT 420  
QY 421 KECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480  
Db 421 KECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480  
QY 481 DNDYLILNLHIPGFKPPTSMGTATGSELGRITVFETLCSADCVLFMVDINRKSTNVVE 540  
Db 481 DNDYLILNLHIPGFKPPTSMGTATGSELGRITVFETLCSADCVLFMVDINRKSTNVVE 540  
QY 541 SWGGTKEKOAYTHIIFKNATFTTWFQRTNQDNRFRINDMVKIYSITATNAVGVAS 600  
Db 541 SWGGTKEKOAYTHIIFKNATFTTWFQRTNQDNRFRINDMVKIYSITATNAVGVAS 600  
QY 601 SCRACALGSEQSGSSCVPCPPGHYIEKETNQCKECPDPTYLSIHQVYGKEACIPCGPGSK 660  
Db 601 SCRACALGSEQSGSSCVPCPPGHYIEKETNQCKECPDPTYLSIHQVYGKEACIPCGPGSK 660  
QY 661 NNQDHSVCSDCFFYHEKENQILHYDFSNLSSVGLMNGPSTFKGTYFHHFNISLCGH 720  
Db 661 NNQDHSVCSDCFFYHEKENQILHYDFSNLSSVGLMNGPSTFKGTYFHHFNISLCGH 720  
QY 721 EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFVCQSTIIPSESKGFRAALSSQSIILA 780  
Db 721 EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFVCQSTIIPSESKGFRAALSSQSIILA 780  
QY 781 DTFIGTVETTLKNNINIKEDMFPVPVPTSOIPDVHFFYKSTATTSCINGRSTAVKMRCNPT 840  
Db 781 DTFIGTVETTLKNNINIKEDMFPVPVPTSOIPDVHFFYKSTATTSCINGRSTAVKMRCNPT 840  
QY 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAECPLCTEHDFFEIEGACKRGFQETLYVWN 900  
Db 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAECPLCTEHDFFEIEGACKRGFQETLYVWN 900  
QY 901 EPKWCIKGISLPEKKLATCETVDFWLKVGVAGVGAFATVALLVALTCYFWKKNQK 953  
Db 901 EPKWCIKGISLPEKKLATCETVDFWLKVGVAGVGAFATVALLVALTCYFWKKNQK 953

RESULT 3  
AAB48377  
ID AAB48377 standard; Protein; 464 AA.  
XX  
AC AAB48377;  
XX

DT 20-APR-2001 (first entry)  
XX

DE Human SEC10 protein sequence (clone ID 1795045.0.77).  
XX SECX; cytostatic; gynecological; gene therapy; screening assay; human;  
KW SEC10; chromosomal mapping; forensic biology; cell proliferation; cancer;  
KW cell differentiation; immune associated disorder; gestational disease.  
XX Homo sapiens.  
OS  
XX WO200078802-A2.  
PN  
XX 28-DEC-2000.  
PD  
XX 23-JUN-2000; 2000WO-US17328.  
PF  
XX 23-JUN-1999; 99US-0140584.  
PR 20-JUL-1999; 99US-0144722.  
PR 16-SEP-1999; 99US-0154520.  
PR 22-JUN-2000; 2000US-0604286.  
XX (CURA-) CURAGEN CORP.  
PA  
XX Shimkets RA, Fernandes E, Vernet C, Yang M, Boldog FL;  
PI Herrmann JL;  
PI  
XX WPI; 2001-071385/08.  
DR N-PSDB; AAC84891.  
DR  
XX Polynucleotides encoding SECX proteins useful for treating disease  
PT characterized by an aberrant level of cell proliferation and/or  
PT differentiation like cancer or immune associated disorders -  
XX  
PS Claim 1; Fig 10; 132pp; English.  
XX  
CC The invention relates to human SECX polypeptides and polynucleotides  
CC encoding them. The SECX polypeptides can be expressed by standard  
CC recombinant methodology. The SECX polypeptides are useful for treating  
CC or preventing a SECX-associated disorder. The invention is useful in  
CC screening assays; detection assays (e.g. chromosomal mapping, cell and  
CC tissue typing, forensic biology); predictive medicine (diagnostic assays,  
CC prognostic assays, monitoring clinical trials, and pharmacogenomics); and  
CC methods of treatment (e.g. therapeutic and prophylactic), especially  
CC disorders characterized by aberrant cell proliferation and/or  
CC differentiation like cancer or immune associated disorders or gestational  
CC disease. The present sequence represents a SEC10 protein.  
XX  
SQ Sequence 464 AA;  
  
Query Match 26.4%; Score 271; DB 22; Length 464;  
Best Local Similarity 99.6%; Pred. No. 4.6e-266;  
Matches 451; Conservative 0; Mismatches 0; Indels 2; Gaps 1;  
  
QY 115 MKNQVCSKCGEGTSLGSGIKFDEWDELPAFESNIATFMDTVVGPSDSRDPGCNNSSWIP 174  
Db 1 MKNQVCSKCGEGTSLGSGIKFDEWDELPAFESNIATFMDTVVGPSDSRDPGCNNSSWIP 60  
QY 175 RGNIESNRDDCTVSLIYAVHLKKSGYVFFEQYVDNNIFFEQQNDQCQEMDTTDKW 234  
Db 61 RGNIESNRDDCTVSLIYAVHLKKSGYVFFEQYVDNNIFFEQQNDQCQEMDTTDKW 120  
QY 235 VKLTDNGEWSHVMKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFFCK 294  
Db 121 VKLTDNGEWSHVMKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFFCK 180  
QY 295 PGTFSENKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQI 352  
Db 181 PGTFSENKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQI 240  
QY 353 HTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPNGFYNNSSSCHPCP 412  
Db 241 HTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPNGFYNNSSSCHPCP 300  
QY 413 PGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 472  
XX



Db 301 PGTFSDGTEKCRPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 360

Qy 473 IQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITFVFETLCSADCVLYFMVDIN 532

Db 361 IQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITFVFETLCSADCVLYFMVDIN 420

Qy 533 RKSTNVVSWGGTKEKQAYTHIIFKNATFTFTW 565

Db 421 RKSTNVVSWGGTKEKQAYTHIIFKNATFTFTW 453

RESULT 4

AAB48372

ID AAB48372 standard; Protein; 411 AA.

XX

AC AAB48372;

DT 20-APR-2001 (first entry)

XX

DE Human SEC5 protein sequence (clone ID 1795045.0.61).

XX

KW SECX; cytostatic; gynecological; gene therapy; screening assay; human;

KW SEC5; chromosomal mapping; forensic biology; cell proliferation; cancer;

KW cell differentiation; immune associated disorder; gestational disease.

XX

OS Homo sapiens.

XX

PN WO200078802-A2.

XX

PD 28-DEC-2000.

XX

PF 23-JUN-2000; 2000WO-US17328.

XX

PR 23-JUN-1999; 99US-0140584.

PR 20-JUL-1999; 99US-0144722.

PR 16-SEP-1999; 99US-0154520.

PR 22-JUN-2000; 2000US-0604286.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Shimkets RA, Fernandes E, Vernet C, Yang M, Boldog FL;

PI Herrmann JL;

XX

DR WPI; 2001-071385/08.

DR N-PSDB; AAC84886.

XX

PT Polynucleotides encoding SECX proteins useful for treating disease

PT characterized by an aberrant level of cell proliferation and/or

PT differentiation like cancer or immune associated disorders -

XX

PS Claim 1; Fig 6; 132pp; English.

XX

CC The invention relates to human SECX polypeptides and polynucleotides

CC encoding them. The SECX polypeptides can be expressed by standard

CC recombinant methodology. The SECX polypeptides are useful for treating

CC or preventing a SECX-associated disorder. The invention is useful in

CC screening assays; detection assays (e.g. chromosomal mapping, cell and

CC tissue typing, forensic biology); predictive medicine (diagnostic assays,

CC prognostic assays, monitoring clinical trials, and pharmacogenomics); and

CC methods of treatment (e.g. therapeutic and prophylactic), especially

CC disorders characterized by aberrant cell proliferation and/or or

CC differentiation like cancer or immune associated disorders or gestational

CC disease. The present sequence represents a SEC5 protein.

XX

SQ Sequence 411 AA;

Query Match 22.6%; Score 232; DB 22; Length 411;

Best Local Similarity 100.0%; Pred. No. 1.7e-226;

Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 334 GSSECTERPPCTTKDYFQIHPTPCDEBEGKTQIMYKWIPEKICREDLTDAILPPSGEKKDC 393

Db 169 GSSECTERPPCTTKDYFQIHPTPCDEBEGKTQIMYKWIPEKICREDLTDAILPPSGEKKDC 228

Qy 394 PPCNPGFYNNGSSSCHPCPPGTFSDGTEKCRPCPAGTEPALGFEYKWNVLPGNMKTSCF 453

Db 229 PPCNPGFYNNGSSSCHPCPPGTFSDGTEKCRPCPAGTEPALGFEYKWNVLPGNMKTSCF 288

Qy 454 NVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITF 513

Db 289 NVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITF 348

Qy 514 VFETLCSADCVLYFMVDINRKSTNVVSWGGTKEKQAYTHIIFKNATFTFTW 565

Db 349 VFETLCSADCVLYFMVDINRKSTNVVSWGGTKEKQAYTHIIFKNATFTFTW 400

RESULT 5

ABB39918

ID ABB39918 standard; Peptide; 81 AA.

XX

AC ABB39918;

XX

DT 04-FEB-2002 (first entry)

XX

DE Peptide #7424 encoded by human foetal liver single exon probe.

XX

KW Human; foetal liver; gene expression; single exon nucleic acid probe.

XX

OS Homo sapiens.

XX

PN WO200157277-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US00669.

XX

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2001-483447/52.

XX

PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human fetal liver -

XX

PS Claim 27; SEQ ID NO 32553; 639pp + sequence listing; English.

XX

CC The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal

CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human

CC foetal liver. The present sequence is a peptide encoded by a single exon

CC nucleic acid probe of the invention.

CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 81 AA;

Query Match 7.9%; Score 81; DB 22; Length 81;

Best Local Similarity 100.0%; Pred. No. 9.8e-74;

Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 578 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNQCKECP 637

Db 1 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNQCKECP 60



```
QY 638 DTYLSIHQVYGKEACIPCGPG 658
Db 61 DTYLSIHQVYGKEACIPCGPG 81

RESULT 6
ABB24471
ID ABB24471 standard; Protein; 81 AA.
XX
AC ABB24471;
XX
DT 23-JAN-2002 (first entry)
XX
DE Protein #6470 encoded by probe for measuring heart cell gene expression.
XX
KW Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 15; SEQ ID No 26241; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 81 AA;

Query Match 7.9%; Score 81; DB 22; Length 81;
Best Local Similarity 100.0%; Pred. No. 9.8e-74;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGGHYIEKETNQCKECP 637
Db 1 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGGHYIEKETNQCKECP 60

QY 638 DTYLSIHQVYGKEACIPCGPG 658
Db 61 DTYLSIHQVYGKEACIPCGPG 81

RESULT 8
AAM73335
ID AAM73335 standard; Protein; 81 AA.
XX
AC AAM73335;
XX
DT 06-NOV-2001 (first entry)
XX
```

```
RESULT 7
AAM60663
ID AAM60663 standard; Protein; 81 AA.
XX
AC AAM60663;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 32768.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 32768; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 81 AA;

Query Match 7.9%; Score 81; DB 22; Length 81;
Best Local Similarity 100.0%; Pred. No. 9.8e-74;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGGHYIEKETNQCKECP 637
Db 1 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGGHYIEKETNQCKECP 60

QY 638 DTYLSIHQVYGKEACIPCGPG 658
Db 61 DTYLSIHQVYGKEACIPCGPG 81

RESULT 8
AAM73335
ID AAM73335 standard; Protein; 81 AA.
XX
AC AAM73335;
XX
DT 06-NOV-2001 (first entry)
XX
```

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 33641.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00668.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488900/53.  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -  
XX  
PS Example 4; SEQ ID NO: 33641; 658pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention.  
XX  
SQ Sequence 81 AA;  
  
Query Match 7.9%; Score 81; DB 22; Length 81;  
Best Local Similarity 100.0%; Pred. No. 9.8e-74;  
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 578 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNQCKECP 637  
Db 1 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNQCKECP 60  
  
QY 638 DTYLSIHQVYGKEACIPCGPG 658  
Db 61 DTYLSIHQVYGKEACIPCGPG 81  
  
RESULT 9  
AAM33535  
ID AAM33535 standard; Protein; 81 AA.  
XX  
AC AAM33535;  
XX  
DT 17-OCT-2001 (first entry)  
XX  
DE Peptide #7572 encoded by probe for measuring placental gene expression.  
XX  
KW probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200157272-A2.  
XX  
PD 09-AUG-2001.  
XX

PE 30-JAN-2001; 2001WO-US00663.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-48897/53.  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human placenta -  
XX  
PS Claim 27; SEQ ID NO 33804; 654pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes (SENP;  
CC see AAT13115-AAI57546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders.  
XX  
SQ Sequence 81 AA;  
  
Query Match 7.9%; Score 81; DB 22; Length 81;  
Best Local Similarity 100.0%; Pred. No. 9.8e-74;  
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 578 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNQCKECP 637  
Db 1 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNQCKECP 60  
  
QY 638 DTYLSIHQVYGKEACIPCGPG 658  
Db 61 DTYLSIHQVYGKEACIPCGPG 81  
  
RESULT 10  
ABG43186  
ID ABG43186 standard; Peptide; 81 AA.  
XX  
AC ABG43186;  
XX  
DT 19-AUG-2002 (first entry)  
XX  
DE Human peptide encoded by genome-derived single exon probe SEQ ID 32851.  
XX  
KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
KW primary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW hyaline membrane disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200186003-A2.  
XX  
PD 15-NOV-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00665.  
XX  
PR 04-FEB-2000; 2000US-180312P.  
PR 26-MAY-2000; 2000US-207456P.  
PR



KW apoptosis; inflammatory; cancer; immune; neurodegenerative.  
XX Unidentified.  
OS WO200112671-A1.  
PN XX  
XX PD  
XX 22-FEB-2001.  
XX PF 10-AUG-2000; 2000WO-US21885.  
XX PR 12-AUG-1999; 99US-0148348.  
PR 13-AUG-1999; 99US-0148683.  
PR 13-AUG-1999; 99US-0148870.  
PR 16-AUG-1999; 99US-0148758.  
PR 17-AUG-1999; 99US-0149181.  
PR 18-AUG-1999; 99US-0149453.  
PR 19-AUG-1999; 99US-0149498.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA Ruben SM, Young PE, Baker KP;  
XX WPI; 2001-138754/14.  
XX New nucleic acid molecule encoding a TR16 tumor necrosis factor  
PT receptor polypeptide, useful for the diagnosis and treatment of cancer,  
PT autoimmune disorders and cardiovascular diseases -  
XX Disclosure; Page 81; 286pp; English.  
XX The present invention relates to a TR16 receptor (tumour necrosis  
CC factor receptor superfamily). The invention is useful treating  
CC diseases and disorders associated with the inhibited or increased  
CC apoptosis. In particular inflammatory diseases, cancers, immune and  
CC neurodegenerative disorders may be treated.  
XX SQ Sequence 74 AA;  
Query Match 7.2%; Score 74; DB 22; Length 74;  
Best Local Similarity 100.0%; Pred. No. 1.2e-66;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 338 CTERPCTTKDYFQIHTPCDEEGKTQIMYKWIETKICREDLTDAIRLPPSGEKKDCPPCN 397  
DB 1 CTERPCTTKDYFQIHTPCDEEGKTQIMYKWIETKICREDLTDAIRLPPSGEKKDCPPCN 60  
QY 398 PGFYNNGSSSCHPC 411  
DB 61 PGFYNNGSSSCHPC 74  
RESULT 13  
AAM72925  
ID AAM72925 standard; Protein; 64 AA.  
XX AAM72925;  
AC AAM72925;  
DT 06-NOV-2001 (first entry)  
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 33231.  
DE Human; bone marrow expressed exon; gene expression analysis; probe;  
XX microarray; cancer; leukaemia; lymphoma; myeloma.  
KW Homo sapiens.  
XX WO200157276-A2.  
PN 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US00668.  
PF 04-FEB-2000; 2000US-0180312.  
PR

PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488900/53.  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -  
XX Example 4; SEQ ID NO: 33231; 658pp + Sequence Listing; English.  
XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention.  
XX SQ Sequence 64 AA;  
Query Match 6.2%; Score 64; DB 22; Length 64;  
Best Local Similarity 100.0%; Pred. No. 1.5e-56;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 105 FSCASGEYLEMKNQVCSCGEGYSLGSGIKFDEWDELPAFSGNIATFMDTVVGPSDSRP 164  
DB 1 FSCASGEYLEMKNQVCSCGEGYSLGSGIKFDEWDELPAFSGNIATFMDTVVGPSDSRP 60  
QY 165 DGCN 168  
DB 61 DGCN 64  
RESULT 14  
ABG42760  
ID ABG42760 standard; Peptide; 64 AA.  
XX AC ABG42760;  
XX DT 19-AUG-2002 (first entry)  
XX Human peptide encoded by genome-derived single exon probe SEQ ID 32425.  
DE Human; single exon probe; asthma; lung cancer; COPD; ILD;  
XX Chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberosus sclerosis; Gaucher's disease; Niemann-pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
KW primary ciliary dyskinesia; fibrocystic pulmonary dysplasia;  
KW hyaline membrane disease.  
XX Homo sapiens.  
OS WO200186003-A2.  
XX PN 15-NOV-2001.  
XX PD 30-JAN-2001; 2001WO-US00665.  
XX 04-FEB-2000; 2000US-180312P.  
PR 26-MAY-2000; 2000US-207456P.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-234687P.







GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.  
OM protein - protein search, using sw model  
Run on: May 12, 2003, 13:23:08 ; Search time 33.8744 Seconds  
(without alignments)  
3788.125 Million cell updates/sec  
Title: US-10-073-333A-2  
Perfect score: 963  
Sequence: 1 MLFRAGPVRGRGWGRPAEA.....TCYFWKKKQKKKTILNLFN 963  
Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0  
Searched: 908470 seqs, 133250620 residues  
Word size : 0  
Total number of hits satisfying chosen parameters: 908470  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Listing first 45 summaries

Database :					A_Geneseq_101002.*				
1:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT.*				
2:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT.*				
3:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT.*				
4:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT.*				
5:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT.*				
6:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT.*				
7:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT.*				
8:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT.*				
9:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT.*				
10:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT.*				
11:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT.*				
12:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT.*				
13:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT.*				
14:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT.*				
15:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT.*				
16:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT.*				
17:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT.*				
18:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT.*				
19:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT.*				
20:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT.*				
21:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT.*				
22:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT.*				
23:					/SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT.*				
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.									
SUMMARIES									
Result No.	Score	Query Match	Length	ID	Description				
1	963	100.0	963	22	AA070255	TR16-short recepto			
2	953	99.0	1027	22	AA070256	TR16-long receptor			
3	271	28.1	464	22	AA048377	Human SEC10 protei			
4	232	24.1	411	22	AA048372	Human SEC5 protein			
5	81	8.4	81	22	AA039918	Peptide #7424 enco			
6	81	8.4	81	22	AA024471	Protein #6470 enco			
7	81	8.4	81	22	AA060663	Human brain expres			
8	81	8.4	81	22	AA073335	Human bone marrow			
9	81	8.4	81	22	AA033535	Peptide #7572 enco			
10	81	8.4	81	23	AA043186	Human peptide enco			

11	74	7.7	74	22	AA070281	Peptide #25. Unid
12	64	6.6	64	22	AA072925	Human bone marrow
13	64	6.6	64	23	AA042760	Human peptide enco
14	60	6.2	60	22	AA038686	Peptide #6192 enco
15	60	6.2	60	22	AA023759	Protein #5758 enco
16	60	6.2	60	22	AA059318	Human brain expres
17	60	6.2	60	22	AA071867	Human bone marrow
18	60	6.2	60	22	AA032149	Peptide #6186 enco
19	60	6.2	60	23	AA041680	Human peptide enco
20	15	1.6	372	22	AA085768	Human seven-transm
21	15	1.6	383	22	AA083853	Amino acid sequenc
22	15	1.6	870	22	AA083851	Amino acid sequenc
23	15	1.6	911	22	AA083850	Amino acid sequenc
24	15	1.6	1001	22	AA035333	Human TR13 recepto
25	15	1.6	1013	21	AA026179	Human CAS8619 prot
26	15	1.6	1013	22	AA012190	Human PRO4985 poly
27	15	1.6	1013	22	AA083845	Amino acid sequenc
28	11	1.1	52	22	AA083848	Peptide fragment o
29	11	1.1	105	21	AA026180	Human CAS8619 prot
30	11	1.1	209	22	AA083852	Amino acid sequenc
31	11	1.1	495	20	AA059972	Human endometrium
32	11	1.1	750	22	AA035328	Human TR13 recepto
33	9	0.9	212	22	AA048483	Novel human diagno
34	9	0.9	708	22	AA091270	C glutamicum prote
35	9	0.9	1487	19	AA076719	Thermococcus 9N2 b
36	9	0.9	1487	19	AA076720	Thermococcus 9N2 m
37	8	0.8	8	22	AA070257	Peptide #1. Unid
38	8	0.8	8	22	AA070258	Peptide #2. Unid
39	8	0.8	8	22	AA070259	Peptide #3. Unid
40	8	0.8	8	22	AA070260	Peptide #4. Unid
41	8	0.8	8	22	AA070261	Peptide #5. Unid
42	8	0.8	8	22	AA070262	Peptide #6. Unid
43	8	0.8	8	22	AA070263	Peptide #7. Unid
44	8	0.8	8	22	AA070264	Peptide #8. Unid
45	8	0.8	8	22	AA070265	Peptide #9. Unid

ALIGNMENTS

RESULT 1	
AA070255	AA070255 standard; protein; 963 AA.
ID	AA070255 standard; protein; 963 AA.
XX	
AC	AA070255;
XX	
DT	10-MAY-2001 (first entry)
XX	
DE	TR16-short receptor protein.
XX	
KW	TR16 receptor; tumour necrosis factor receptor superfamily;
KW	apoptosis; inflammatory; cancer; immune; neurodegenerative.
XX	
OS	Unidentified.
XX	
PN	WO200112671-A1.
XX	
PD	22-FEB-2001.
XX	
PF	10-AUG-2000; 2000WO-US21885.
XX	
PR	12-AUG-1999; 99US-0148348.
PR	13-AUG-1999; 99US-0148683.
PR	13-AUG-1999; 99US-0148870.
PR	16-AUG-1999; 99US-0148758.
PR	17-AUG-1999; 99US-0149181.
PR	18-AUG-1999; 99US-0149453.
PR	19-AUG-1999; 99US-0149498.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Ruben SM, Young PE, Baker KP;
XX	





QY	61	CQEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLDPPVRGKECTFSCASGEYLEMKNQVC	120
Db	61		
QY	121	SKCGEGTYSLGSIGIKFDEWDELPAAGFSNIATFMDTVVGPSPDSRDPDGCNNSSWI	180
Db	121		
QY	181	SNRDDCTVSLIYAVHLKKSQYVFEFYQYVDNNIFFEFTQNDQCQEMD	240
Db	181		
QY	241	GEWGSYMLKSGTNIYWRRTGILMGSKAVKPVLVKNITIEGVAYTSECPCPKGTF	300
Db	241		
QY	301	KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG	360
Db	301		
QY	361	KTQIMYKWIPIKICREDLTDAIRLPPSGEKKDCPPCNPNGFYNNSSSCHPCPPGTFSDGT	420
Db	361		
QY	421	KECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS	480
Db	421		
QY	481	DNDYLILNLHIPGFKPPTSMGTATGSELGRITFVFETLCSADCVLFMVDINRKSTNVVE	540
Db	481		
QY	541	SWGKTKEKQAYTHIFKNATFTTFAQRTNQGDNRREINDMVKIYSITATNAVDG	600
Db	541		
QY	601	SCRACALGSEQSGSCVPCPPGHYIEKETNQCKECPDPTLSIHQVYKACIPCGPGSK	660
Db	601		
QY	661	NNQDHSVCYSDCFFYHEKENQILHYDFSNLSSVGLMNGPSTSKGTYFHFNFNISLCGH	720
Db	661		
QY	721	EGKKNALCTNNITDFTVKEIVAGSDDYTNLVGAFVCQSTIIPSESKGFRAALSSQSIILA	780
Db	721		
QY	781	DTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSATTSCINGRSTAVKMR	840
Db	781		
QY	841	KSGAGVISVPSKCPAGTCDGCTFFFLWESAEACPLCTEHDFEIEGACKRGFQETLYV	900
Db	841		
QY	901	EPKWCIGISLPEKKLATCETVDFWLKVGVAGVGAFTAVLLVALTCTCYFWKKNOK	953
Db	901		
RESULT 3			
AAB48377			
ID	AAB48377 standard; Protein; 464 AA.		
XX			
AC	AAB48377;		
XX	20-APR-2001 (first entry)		
DE	Human SEC10 protein sequence (clone ID 1795045.0.77).		
XX	SECX; cytostatic; gynecological; gene therapy; screening assay; human;		
KW	SEC10; chromosomal mapping; forensic biology; cell proliferation; cancer;		

KW	cell differentiation; immune associated disorder; gestational disease.
XX	
OS	Homo sapiens.
XX	
PN	WO200078802-A2.
XX	
PD	28-DEC-2000.
XX	
PF	23-JUN-2000; 2000WO-US17328.
XX	
PR	23-JUN-1999; 99US-0140584.
PR	20-JUL-1999; 99US-0144722.
PR	16-SEP-1999; 99US-0154520.
PR	22-JUN-2000; 2000US-0604286.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shimkets RA, Fernandes E, Vernet C, Yang M, Boldog FL;
PI	Herrmann JL;
XX	
DR	WPI; 2001-071385/08.
XX	N-PSDB; AAC84891.
PT	Polynucleotides encoding SECX proteins useful for treating disease
PT	characterized by an aberrant level of cell proliferation and/or
PT	differentiation like cancer or immune associated disorders.
XX	
PS	Claim 1; Fig 10; 132pp; English.
XX	
CC	The invention relates to human SECX polypeptides and polynucleotides
CC	encoding them. The SECX polypeptides can be expressed by standard
CC	recombinant methodology. The SECX polypeptides are useful for treating
CC	or preventing a SECX-associated disorder. The invention is useful in
CC	screening assays; detection assays (e.g. chromosomal mapping, cell and
CC	tissue typing, forensic biology); predictive medicine (diagnostic assays,
CC	prognostic assays, monitoring clinical trials, and pharmacogenomics); and
CC	methods of treatment (e.g. therapeutic and prophylactic), especially
CC	disorders characterized by aberrant cell proliferation and/or
CC	differentiation like cancer or immune associated disorders or gestational
CC	disease. The present sequence represents a SEC10 protein.
XX	
SQ	Sequence 464 AA;

Query Match

28.1%; Score 271; DB 22; Length 464;

Best Local Similarity

99.6%; Pred. No. 1.7e-264;

Matches 451; Conservative

0; Mismatches 0; Indels 2; Gaps 1;

QY	115	MKNQVCSKCGEGTYSLGSGIKFDEWDELPAAGFSNIATFMDTVVGPSPDSRDPG	CNNSSWIP	174
Db	1			
QY	175	RGNYESNRDDCTVSLIYAVHLKKSQYVFEFYQYVDNNIFFEFTQNDQCQEMD	TTTDKW	234
Db	61			
QY	235	VKLTDNGEWGSVMLKSGTNIYWRRTGILMGSKAVKPVLVKNITIEGVAYTSEC	FPCK	294
Db	121			
QY	295	PGTFSNKPQSPNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTERPPCT	TKDYFQI	352
Db	181			
QY	353	HTPCDEEGKTQIMYKWIPIKICREDLTDAIRLPPSGEKKDCPPCNPNGFYNN	SSSCHPCP	412
Db	241			
QY	413	PGTFSNKPQSPNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTERPPCT	TKDYFQI	240
Db	301			
QY	473	IQSGAGGSDNDYLILNLHIPGFKPPTSMGTATGSELGRITFVFETLCSADCV	LFYMVDIN	532

Db 361 IQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITFVFETLCSADCVLFMVDIN 420

QY 533 RKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 565  
|||||

Db 421 RKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 453

RESULT 4  
AAB48372  
ID AAB48372 standard; Protein; 411 AA.  
XX AAB48372;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE Human SEC5 protein sequence (clone ID 1795045.0.61).  
XX  
KW SECX; cytostatic; gynecological; gene therapy; screening assay; human;  
KW SEC5; chromosomal mapping; forensic biology; cell proliferation; cancer;  
KW cell differentiation; immune associated disorder; gestational disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200078802-A2.  
XX  
PD 28-DEC-2000.  
XX  
PF 23-JUN-2000; 2000WO-US17328.  
XX  
PR 23-JUN-1999; 99US-0140584.  
PR 20-JUL-1999; 99US-0144722.  
PR 16-SEP-1999; 99US-0154520.  
PR 22-JUN-2000; 2000US-0604286.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shimkets RA, Fernandes E, Vernet C, Yang M, Boldog FL;  
PI Herrmann JL;  
XX  
DR WPI; 2001-071385/08.  
DR N-PSDB; AAC84886.  
XX  
PT Polynucleotides encoding SECX proteins useful for treating disease  
PT characterized by an aberrant level of cell proliferation and/or  
PT differentiation like cancer or immune associated disorders -  
XX  
PS Claim 1; Fig 6; 132pp; English.  
XX  
CC The invention relates to human SECX polypeptides and polynucleotides  
CC encoding them. The SECX polypeptides can be expressed by standard  
CC recombinant methodology. The SECX polypeptides are useful for treating  
CC or preventing a SECX-associated disorder. The invention is useful in  
CC screening assays; detection assays (e.g. chromosomal mapping, cell and  
CC tissue typing, forensic biology); predictive medicine (diagnostic assays,  
CC prognostic assays, monitoring clinical trials, and pharmacogenomics); and  
CC methods of treatment (e.g. therapeutic and prophylactic), especially  
CC disorders characterized by aberrant cell proliferation and/ or  
CC differentiation like cancer or immune associated disorders or gestational  
CC disease. The present sequence represents a SEC5 protein.  
XX  
SQ Sequence 411 AA;

Query Match 24.1%; Score 232; DB 22; Length 411;  
Best Local Similarity 100.0%; Pred. No. 3.7e-225;  
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 GSSECTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIETKICREDLTDAILRPPSGEKKDC 393  
|||||

Db 169 GSSECTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIETKICREDLTDAILRPPSGEKKDC 228

QY 394 PPCNPGFYNNSSSCHPCPPGTGTSKTCRCPAGTEPALGFYKWNVLPGNMKTSKF 453  
|||||

Db 229 PPCNPGFYNNSSSCHPCPPGTGTSKTCRCPAGTEPALGFYKWNVLPGNMKTSKF 288

QY 454 NVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITF 513  
|||||

Db 289 NVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITF 348  
|||||

QY 514 VFETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 565  
|||||

Db 349 VFETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 400  
|||||

RESULT 5  
ABB39918  
ID ABB39918 standard; Peptide; 81 AA.  
XX  
AC ABB39918;  
XX  
DT 04-FEB-2002 (first entry)  
XX  
DE Peptide #7424 encoded by human foetal liver single exon probe.  
XX  
KW Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX  
OS Homo sapiens.  
XX  
PN WO200157277-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00669.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-483447/52.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human fetal liver -  
XX  
PS Claim 27; SEQ ID NO 32553; 639pp + sequence listing; English.  
XX  
CC The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a peptide encoded by a single exon  
CC nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 81 AA;

Query Match 8.4%; Score 81; DB 22; Length 81;  
Best Local Similarity 100.0%; Pred. No. 2.7e-73;  
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNQCKECP 637  
|||||

Db 1 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNQCKECP 60  
|||||

QY 638 DTYLSIHQVYGKEACIPCGPG 658  
|||||

Db 61 DTYLSIHQVYGKEACIPCGPG 81  
|||||

RESULT 6  
ABB24471  
ID ABB24471 standard; Protein; 81 AA.  
XX AC ABB24471;  
XX AC ABB24471;  
DT 23-JAN-2002 (first entry)  
XX DE Protein #6470 encoded by probe for measuring heart cell gene expression.  
XX OS Homo sapiens.  
KW Human; gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.

XX OS Homo sapiens.  
XX WO200157274-A2.  
XX PN 09-AUG-2001.  
XX PD 30-JAN-2001; 2001WO-US00666.  
XX PF 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488899/53.  
XX Single exon nucleic acid probes for analyzing gene expression in human hearts -  
XX Claim 15; SEQ ID NO 26241; 530pp; English.  
XX The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart (see ABA21535-ABA41305). The present sequence is a protein encoded by one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays.

XX CC By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.  
XX CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 81 AA;  
Query Match 8.4%; Score 81; DB 22; Length 81;  
Best Local Similarity 100.0%; Pred. No. 2.7e-73;  
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 RFINDMWKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNOCKECP 637  
Db 1 RFINDMWKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNOCKECP 60  
QY 638 DTYLSIHQVYGKEACIPCGPG 658  
Db 61 DTYLSIHQVYGKEACIPCGPG 81

RESULT 7  
AAM60663  
ID AAM60663 standard; Protein; 81 AA.

XX AC AAM60663;  
XX DT 05-NOV-2001 (first entry)  
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 32768.  
XX KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer.

XX OS Homo sapiens.  
XX PN WO200157275-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00667.  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX Single exon nucleic acid probes for analyzing gene expression in human brains -  
XX Example 4; SEQ ID NO: 32768; 650pp + Sequence Listing; English.

XX CC The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is a protein encoded by one of the probes of the invention.

XX SQ Sequence 81 AA;  
Query Match 8.4%; Score 81; DB 22; Length 81;  
Best Local Similarity 100.0%; Pred. No. 2.7e-73;  
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 RFINDMWKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNOCKECP 637  
Db 1 RFINDMWKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNOCKECP 60  
QY 638 DTYLSIHQVYGKEACIPCGPG 658  
Db 61 DTYLSIHQVYGKEACIPCGPG 81

RESULT 8  
AAM73335  
ID AAM73335 standard; Protein; 81 AA.  
XX AC AAM73335;  
XX DT 06-NOV-2001 (first entry)  
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 33641.  
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.

```
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX XX 09-AUG-2001.
XX PD
XX PF 30-JAN-2001; 2001WO-US00668.
XX XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX
XX DR WPI; 2001-488900/53.
XX XX
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow -
XX XX
XX PS Example 4; SEQ ID NO: 33641; 658pp + Sequence Listing; English.
XX XX
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a
XX CC protein encoded by one of the probes of the invention.
XX XX
XX SQ Sequence 81 AA;

Query Match 8.4%; Score 81; DB 22; Length 81;
Best Local Similarity 100.0%; Pred. No. 2.7e-73;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGHHYIEKETNQCKECP 637
Db 1 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGHHYIEKETNQCKECP 60

QY 638 DTYLSIHQVYGKEACIPCGPG 658
Db 61 DTYLSIHQVYGKEACIPCGPG 81

RESULT 9
AAM33535
ID AAM33535 standard; Protein; 81 AA.
XX AC AAM33535;
XX XX
XX DT 17-OCT-2001 (first entry)
XX XX
XX DE Peptide #7572 encoded by probe for measuring placental gene expression.
XX XX
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX XX
XX OS Homo sapiens.
XX XX
XX PN WO200157272-A2.
XX XX
XX PD 09-AUG-2001.
XX XX
XX PF 30-JAN-2001; 2001WO-US00663.
XX XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
```

```
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX
XX DR WPI; 2001-48897/53.
XX XX
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX XX
XX PS Claim 27; SEQ ID No 33804; 654pp; English.
XX XX
XX CC The present invention relates to single exon nucleic acid probes (SENP:
XX CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders.
XX XX
XX SQ Sequence 81 AA;

Query Match 8.4%; Score 81; DB 22; Length 81;
Best Local Similarity 100.0%; Pred. No. 2.7e-73;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGHHYIEKETNQCKECP 637
Db 1 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGHHYIEKETNQCKECP 60

QY 638 DTYLSIHQVYGKEACIPCGPG 658
Db 61 DTYLSIHQVYGKEACIPCGPG 81

RESULT 10
ABG43186
ID ABG43186 standard; Peptide; 81 AA.
XX AC ABG43186;
XX XX
XX DT 19-AUG-2002 (first entry)
XX XX
XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 32851.
XX XX
XX KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX KW chronic obstructive pulmonary disease; interstitial lung disease;
XX KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemoiderosis;
XX KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX KW primary ciliary dyskinesia; pulmonary hypertension;
XX KW hyaline membrane disease.
XX XX
XX OS Homo sapiens.
XX XX
XX PN WO200186003-A2.
XX XX
XX PD 15-NOV-2001.
XX XX
XX PF 30-JAN-2001; 2001WO-US00665.
XX XX
XX PR 04-FEB-2000; 2000US-180312P.
XX PR 26-MAY-2000; 2000US-207456P.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-234687P.
XX PR 27-SEP-2000; 2000US-236359P.
```



PR 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2002-114183/15.  
DR Spatially-addressable set of single exon nucleic acid probes, used to  
XX measure gene expression in human lung samples -  
PT Claim 27; SEQ ID No 32851; 634pp; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of  
CC probes; the novel set of probes which hybridise at high stringency to a  
CC nucleic acid expressed in the human lung; measuring gene expression in a  
CC sample derived from human lung, comprising (a) contacting the array with  
CC a collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of  
CC the array; identifying exons in a eukaryotic genome, comprising  
CC (a) algorithmically predicting at least one exon from genomic sequences  
CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene  
CC expression analysis, and for identifying exons in a gene, particularly  
CC using human lung derived mRNA and for the study of lung diseases  
CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
CC haemosiderosis, pulmonary histiocytosis, lymphangioidomyomatosis,  
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
CC and hyaline membrane disease. The present sequence is a peptide/protein  
CC encoded by a single exon probe of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 81 AA;  
  
Query Match 8.4%; Score 81; DB 23; Length 81;  
Best Local Similarity 100.0%; Pred. No. 2.7e-73;  
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 578 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNCKECP 637  
Db 1 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNCKECP 60  
  
Qy 638 DFVLSIHQVYGKEACIPCGPG 658  
Db 61 DFVLSIHQVYGKEACIPCGPG 81  
  
RESULT 11  
AAB70281  
ID AAB70281 standard; peptide; 74 AA.  
XX

AC AAB70281;  
XX  
DT 10-MAY-2001 (first entry)  
XX  
DE Peptide #25.  
XX  
KW TR16 recptor; tumour necrosis factor receptor superfamily;  
KW apoptosis; inflammatory; cancer; immune; neurodegenerative.  
XX  
OS Unidentified.  
XX  
PN WO200112671-A1.  
XX  
XX 22-FEB-2001.  
XX  
PF 10-AUG-2000; 2000WO-US21885.  
XX  
PR 12-AUG-1999; 99US-0148348.  
PR 13-AUG-1999; 99US-0148683.  
PR 13-AUG-1999; 99US-0148870.  
PR 16-AUG-1999; 99US-0148758.  
PR 17-AUG-1999; 99US-0149181.  
PR 18-AUG-1999; 99US-0149453.  
PR 19-AUG-1999; 99US-0149498.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Young PE, Baker KP;  
XX  
DR WPI; 2001-138754/14.  
XX  
XX New nucleic acid molecule encoding a TR16 tumor necrosis factor  
PT receptor polypeptide, useful for the diagnosis and treatment of cancer,  
PT autoimmune disorders and cardiovascular diseases -  
XX  
PS Disclosure; Page 81; 286pp; English.  
XX  
CC The present invention relates to a TR16 receptor (tumour necrosis  
CC factor receptor superfamily). The invention is useful treating  
CC diseases and disorders associated with the inhibited or increased  
CC apoptosis. In particular inflammatory diseases, cancers, immune and  
CC neurodegenerative disorders may be treated.  
XX  
SQ Sequence 74 AA;  
  
Query Match 7.7%; Score 74; DB 22; Length 74;  
Best Local Similarity 100.0%; Pred. No. 2.9e-66;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 338 CTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIETKICREDLTDAIRLPPSGEKKDCPPCN 397  
Db 1 CTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIETKICREDLTDAIRLPPSGEKKDCPPCN 60  
  
Qy 398 PGFYNNNGSSSCHPC 411  
Db 61 PGFYNNNGSSSCHPC 74  
  
RESULT 12  
AAM72925  
ID AAM72925 standard; Protein; 64 AA.  
XX  
AC AAM72925;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 33231.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX  
OS Homo sapiens.  
XX

PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000668.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488900/53.  
DR  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -  
PT  
XX Example 4; SEQ ID NO: 33231; 658pp + Sequence Listing; English.  
PS  
XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention.  
XX  
SQ Sequence 64 AA;  
  
Query Match 6.6%; Score 64; DB 22; Length 64;  
Best Local Similarity 100.0%; Pred. No. 3.2e-56;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 105 FSCASGEYLEMKNQVCSKCGEGTYSLGSGIKFDEWDELPAAGFSNIATFMDTVVGPDSRP 164  
Db 1 FSCASGEYLEMKNQVCSKCGEGTYSLGSGIKFDEWDELPAAGFSNIATFMDTVVGPDSRP 60  
  
QY 165 DGCN 168  
Db 61 DGCN 64  
  
RESULT 13  
ABG42760  
ID ABG42760 standard; Peptide; 64 AA.  
XX  
AC ABG42760;  
XX  
DT 19-AUG-2002 (first entry)  
DE  
XX Human peptide encoded by genome-derived single exon probe SEQ ID 32425.  
KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karageners syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200186003-A2.  
XX  
PD 15-NOV-2001.  
XX

PF 30-JAN-2001; 2001WO-US000665.  
XX  
PR 04-FEB-2000; 2000US-180312P.  
PR 26-MAY-2000; 2000US-207456P.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-234687P.  
PR 27-SEP-2000; 2000US-236359P.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2002-114183/15.  
DR  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples -  
PT  
XX Claim 27; SEQ ID No 32425; 634pp; English.  
PS  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of  
CC probes; the novel set of probes which hybridise at high stringency to a  
CC nucleic acid expressed in the human lung; measuring gene expression in a  
CC sample derived from human lung, comprising (a) contacting the array with  
CC a collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of  
CC the array; identifying exons in a eukaryotic genome, comprising  
CC (a) algorithmically predicting at least one exon from genomic sequences  
CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene  
CC expression analysis, and for identifying exons in a gene, particularly  
CC using human lung derived mRNA and for the study of lung diseases  
CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
CC pulmonary alveolar proteinosis, Karageners syndrome, fibrocystic  
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
CC and hyaline membrane disease. The present sequence is a peptide/protein  
CC encoded by a single exon probe of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 64 AA;  
  
Query Match 6.6%; Score 64; DB 23; Length 64;  
Best Local Similarity 100.0%; Pred. No. 3.2e-56;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 105 FSCASGEYLEMKNQVCSKCGEGTYSLGSGIKFDEWDELPAAGFSNIATFMDTVVGPDSRP 164  
Db 1 FSCASGEYLEMKNQVCSKCGEGTYSLGSGIKFDEWDELPAAGFSNIATFMDTVVGPDSRP 60  
  
QY 165 DGCN 168

```

Db      61 DGCN 64

RESULT 14
ABB38686
ID  ABB38686 standard; Peptide; 60 AA.
XX
AC  ABB38686;
XX
DT  04-FEB-2002 (first entry)
XX
DE  Peptide #6192 encoded by human foetal liver single exon probe.
XX
KW  Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS  Homo sapiens.
XX
PN  WO200157277-A2.
XX
PD  09-AUG-2001.
XX
PF  30-JAN-2001; 2001WO-US00669.
XX
PR  04-FEB-2000; 2000US-0180312.
PR  26-MAY-2000; 2000US-0207456.
PR  30-JUN-2000; 2000US-0608408.
PR  03-AUG-2000; 2000US-0632366.
PR  21-SEP-2000; 2000US-0234687.
PR  27-SEP-2000; 2000US-0236359.
PR  04-OCT-2000; 2000GB-0024263.
XX
PA  (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI  Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR  WPI; 2001-483447/52.
XX
PT  Human genome-derived single exon nucleic acid probes useful for
    analyzing gene expression in human fetal liver -
XX
PS  Claim 27; SEQ ID NO 31321; 639pp + sequence listing; English.
XX
CC  The invention relates to a single exon nucleic acid probe for
    measuring human gene expression in a sample derived from human foetal
    liver. The single exon nucleic acid probes may be used for predicting,
    measuring and displaying gene expression in samples derived from human
    fetal liver. The present sequence is a peptide encoded by a single exon
    nucleic acid probe of the invention.
XX
CC  Note: The sequence data for this patent did not form part of the
    printed specification, but was obtained in electronic format directly
    from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ  Sequence 60 AA;

Query Match      6.2%; Score 60; DB 22; Length 60;
Best Local Similarity 100.0%; Pred. No. 3.3e-52;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  362 TQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPNGFYNNGSSSCHPCPPGTFSDGK 421
    |||||
DB  1 TQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPNGFYNNGSSSCHPCPPGTFSDGK 60

RESULT 15
ABB23759
ID  ABB23759 standard; Protein; 60 AA.
XX
AC  ABB23759;
XX
DT  23-JAN-2002 (first entry)
XX
DE  Protein #5758 encoded by probe for measuring heart cell gene expression.

```





GenCore version 5.1.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.  
  
OM protein - protein search, using sw model  
  
Run on: May 12, 2003, 13:31:24 ; Search time 16.4533 Seconds  
(without alignments)  
5626.691 Million cell updates/sec

Title: US-10-073-333A-2  
Perfect score: 963  
Sequence: 1 MLFRARGPVRRGWRPAEA.....TCYFWKKNQKKKTILNLFN 963

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	0.8	100	1 S27016	defensin alpha-6 p
2	8	0.8	148	2 E84239	hypothetical prote
3	8	0.8	184	1 QQCVLG	BL1 protein - toma
4	8	0.8	228	2 S65426	pyruvate decarboxy
5	8	0.8	228	2 S65425	pyruvate decarboxy
6	8	0.8	239	2 G87265	conserved hypothet
7	8	0.8	268	2 AB1424	E. coli RpiR trans
8	8	0.8	268	2 AH1797	E. coli RpiR trans
9	8	0.8	271	2 S12783	OX40 antigen precu
10	8	0.8	272	2 I48700	gene ox40 protein
11	8	0.8	281	2 T33466	hypothetical prote
12	8	0.8	286	2 A82929	ATP synthase gamma
13	8	0.8	405	2 S65471	pyruvate decarboxy
14	8	0.8	417	2 S57820	pyruvate decarboxy
15	8	0.8	428	2 AF0241	probable coenzyme
16	8	0.8	444	2 T01548	protein kinase hom
17	8	0.8	463	2 T28748	hypothetical prote
18	8	0.8	585	2 T03252	pyruvate decarboxy
19	8	0.8	603	2 T03295	pyruvate decarboxy
20	8	0.8	622	2 A36915	fructanase - Bacte
21	8	0.8	671	2 T10755	kinesin-related pr
22	8	0.8	742	2 AE2357	hypothetical prote
23	8	0.8	881	2 T01269	serine/threonine-s
24	8	0.8	1206	2 B87247	probable conserved
25	8	0.8	1578	2 AD1512	peptidoglycan boun
26	8	0.8	1582	2 AC1153	adhesin homolog lm
27	8	0.8	2295	2 B71621	probable membrane
28	8	0.8	3102	2 T43291	laminin alpha chai
29	7	0.7	94	2 A64863	hypothetical prote

30	7	0.7	100	1 TNLJSI	trans-activating t
31	7	0.7	100	2 A43998	hypothetical prote
32	7	0.7	109	2 S42599	hypothetical prote
33	7	0.7	122	2 T28199	hypothetical prote
34	7	0.7	134	2 C84023	hypothetical prote
35	7	0.7	134	2 T36365	proline-rich prote
36	7	0.7	135	2 PC2269	cytochrome P450 pr
37	7	0.7	141	2 A27482	vasotocin / neurop
38	7	0.7	147	2 E70746	probable mmps2 pro
39	7	0.7	157	2 AG3560	transcription regu
40	7	0.7	159	2 E29879	vasotocin / neurop
41	7	0.7	173	2 AB3450	invasion protein b
42	7	0.7	179	2 H64472	hypothetical prote
43	7	0.7	180	2 T49530	related to glycine
44	7	0.7	182	2 A64834	fimbrial-like prot
45	7	0.7	182	2 C85620	probable fimbrial-

ALIGNMENTS

RESULT 1  
S27016  
defensin alpha-6 precursor - human  
N;Alternate names: Paneth cell-specific alpha-defensin 6  
C;Species: Homo sapiens (man)  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: S27016  
R;Jones, D.E.; Bevins, C.L.  
FEBS Lett. 315, 187-192, 1993  
A;Title: Defensin-6 mRNA in human Paneth cells: implications for antimicrobial peptid  
A;Reference number: S27016; MUID:93114459; PMID:8417977  
A;Accession: S27016  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-100 <JON>  
A;Cross-references: GB:M98331; NID:g181546; PIDN:AAB59357.1; PID:g181547  
C;Genetics:

A;Gene: GDB:DEFA6; DEF6; HD-6  
A;Cross-references: GDB:136838; OMIM:600471  
A;Map position: 8pter-8p21  
C;Superfamily: mammalian defensin  
C;Keywords: antibiotic; homodimer  
F;1-19/Domain: signal sequence #status predicted <SIG>  
F;20-70/Domain: propeptide #status predicted <PRO>  
F;71-100/Product: defensin alpha-6 #status predicted <MAT>  
F;72-99,74-88,78-98/Disulfide bonds: #status predicted

Query Match 0.8%; Score 8; DB 1; Length 100;  
Best Local Similarity 100.0%; Pred. No. 4.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 936 TAVLLVAL 943  
|||||||  
Db 8 TAVLLVAL 15

RESULT 2  
E84239  
hypothetical protein Vng0825c [imported] - Halobacterium sp. NRC-1  
C;Species: Halobacterium sp. NRC-1  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: E84239  
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky  
; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;  
A;Title: Genome sequence of Halobacterium species NRC-1.  
A;Reference number: A84160; MUID:20504483; PMID:11016950  
A;Accession: E84239  
A;Status: preliminary  
A;Molecule type: DNA

A;Residues: 1-148 <STO>  
A;Cross-references: GB:AE004437; NID:gl0580395; PIDN:AAG19281.1; GSPDB:GN00138  
C;Genetics:  
A;Gene: VNG0825C

Query Match 0.8%; Score 8; DB 2; Length 148;  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 TVVGPSDS 162  
Db 5 TVVGPSDS 12  
IIIIIIII

RESULT 3  
QQCVLG  
BL1 protein - tomato golden mosaic virus  
C;Species: tomato golden mosaic virus  
A;Note: host Nicotiana sp. (tobacco)  
C;Date: 28-Aug-1985 #sequence\_revision 28-Aug-1985 #text\_change 08-Apr-1994  
C;Accession: A04169  
R;Hamilton, W.D.O.; Stein, V.E.; Coutts, R.H.A.; Buck, K.W.  
EMBO J. 3, 2197-2205, 1984  
A;Title: Complete nucleotide sequence of the infectious cloned DNA components of tomato  
A;Reference number: A04163  
A;Accession: A04169  
A;Molecule type: DNA  
A;Residues: 1-184 <HAM>  
C;Comment: The genome consists of two circular, single-stranded DNA components, DNA A and  
C;Genetics:  
A;Map position: segment B  
C;Superfamily: tomato golden mosaic virus BL1 protein

Query Match 0.8%; Score 8; DB 1; Length 184;  
Best Local Similarity 100.0%; Pred. No. 7.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 NYIESNRD 184  
Db 13 NYIESNRD 20  
IIIIIIII

RESULT 4  
S65426  
pyruvate decarboxylase (EC 4.1.1.1) - fava bean (fragment)  
C;Species: Vicia faba (fava bean)  
C;Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 15-Oct-1999  
C;Accession: S65426  
R;Muecke, U.; Wohlfarth, T.; Fiedler, U.; Baeumlein, H.; Ruecknagel, K.P.; Koenig, S.  
Eur. J. Biochem. 237, 373-382, 1996  
A;Title: Pyruvate decarboxylase from Pisum sativum. Properties, nucleotide and amino acid  
A;Reference number: S65423; MUID:96215432; PMID:8647075  
A;Accession: S65426  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-228 <MUE>  
C;Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain  
C;Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 0.8%; Score 8; DB 2; Length 228;  
Best Local Similarity 100.0%; Pred. No. 9.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLY 276  
Db 50 KAVKPVLY 57  
IIIIIIII

RESULT 5  
S65425  
pyruvate decarboxylase (EC 4.1.1.1) (clone PDC3) - garden pea (fragment)  
C;Species: Pisum sativum (garden pea)

C;Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 15-Oct-1999  
C;Accession: S65425  
R;Muecke, U.; Wohlfarth, T.; Fiedler, U.; Baeumlein, H.; Ruecknagel, K.P.; Koenig, S.  
Eur. J. Biochem. 237, 373-382, 1996  
A;Title: Pyruvate decarboxylase from Pisum sativum. Properties, nucleotide and amino  
A;Reference number: S65423; MUID:96215432; PMID:8647075  
A;Accession: S65425  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-228 <MUE>  
C;Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain  
C;Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 0.8%; Score 8; DB 2; Length 228;  
Best Local Similarity 100.0%; Pred. No. 9.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLY 276  
Db 50 KAVKPVLY 57  
IIIIIIII

RESULT 6  
G87265  
conserved hypothetical protein CC0136 [imported] - Caulobacter crescentus  
C;Species: Caulobacter crescentus  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C;Accession: G87265  
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A;Title: Complete Genome Sequence of Caulobacter crescentus.  
A;Reference number: A87249; MUID:21173698; PMID:11259647  
A;Accession: G87265  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-239 <STO>  
A;Cross-references: GB:AE005673; NID:gl3421247; PIDN:AAK22123.1; GSPDB:GN00148  
C;Genetics:  
A;Gene: CC0136

Query Match 0.8%; Score 8; DB 2; Length 239;  
Best Local Similarity 100.0%; Pred. No. 9.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 AEAPRRGR 25  
Db 176 AEAPRRGR 183  
IIIIIIII

RESULT 7  
AB1424  
E. coli RpiR transcription regulator homolog lmo2795 [imported] - Listeria monocytoge  
C;Species: Listeria monocytogenes  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AB1424  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maltournam, A.;  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A;Title: Comparative-genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AB1424  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-268 <GLA>  
A;Cross-references: GB:NC\_003210; PIDN:CAD01008.1; PID:gl6412295; GSPDB:GN00177  
A;Experimental source: strain EGD-e  
C;Genetics:  
A;Gene: lmo2795

Query Match 0.8%; Score 8; DB 2; Length 268;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 792 LKNINIKE 799  
Db 32 LKNINIKE 39  
|||||

RESULT 8  
AH1797  
E. coli RpiR transcription regulator homolog lin2927 [imported] - Listeria innocua (stra  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AH1797  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AH1797  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-268 <GLA>  
A;Cross-references: GB:AL592022; PIDN:CAC98152.1; PID:g16415468; GSPDB:GN00178  
A;Experimental source: strain Clip11262  
C;Genetics:  
A;Gene: lin2927

Query Match 0.8%; Score 8; DB 2; Length 268;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 792 LKNINIKE 799  
Db 32 LKNINIKE 39  
|||||

RESULT 9  
S12783  
OX40 antigen precursor - rat  
N;Alternate names: nerve growth factor receptor homolog  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 05-Nov-1999  
C;Accession: S12783; S08036  
R;Mallett, S.; Fossum, S.; Barclay, A.N.  
EMBO J. 9, 1063-1068, 1990  
A;Title: Characterization of the MRC OX40 antigen of activated CD4 positive T lymphocyte  
A;Reference number: S12783; MUID:90214614; PMID:2157591  
A;Accession: S12783  
A;Molecule type: mRNA  
A;Residues: 1-271 <MAL>  
A;Cross-references: EMBL:X17037; NID:g57830; PIDN:CAA34897.1; PID:g57831  
C;Superfamily: CD27 antigen; NGF receptor repeat homology  
C;Keywords: growth factor receptor; transmembrane protein  
F;1-19/Domain: signal sequence #status predicted <SIG>  
F;20-271/Product: OX40 antigen #status predicted <MAT>  
F;211-235/Domain: transmembrane #status predicted <TM>

Query Match 0.8%; Score 8; DB 2; Length 271;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623  
Db 122 CVPCPPGH 129  
|||||

RESULT 10

I48700  
gene ox40 protein - mouse  
N;Alternate names: OX40 antigen  
C;Species: Mus musculus (house mouse)  
C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 11-Jan-2000  
C;Accession: I48700; I48334; S34377  
R;Calderhead, D.M.; Buhlmann, J.E.; van den Eertwegh, A.J.; Claassen, E.; Noelle, R.J  
J. Immunol. 151, 5261-5271, 1993  
A;Title: Cloning of mouse Ox40: a T cell activation marker that may mediate T-B cell  
A;Reference number: I48700; MUID:94044750; PMID:8228223  
A;Accession: I48700  
A;Status: translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-272 <RES>  
A;Cross-references: EMBL:Z21674; NID:g312827; PIDN:CAA79772.1; PID:g312828  
R;Birkeland, M.L.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.; Barclay, A.N.  
Eur. J. Immunol. 25, 926-930, 1995  
A;Title: Gene structure and chromosomal localization of the mouse homologue of rat OX  
A;Reference number: I48334; MUID:95255413; PMID:7737295  
A;Accession: I48334  
A;Status: translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-14,'G',16-272 <RE2>  
A;Cross-references: EMBL:X85214; NID:g732818; PIDN:CAA59476.1; PID:g732819  
C;Genetics:  
A;Gene: Ox40  
A;Introns: 45/1; 86/1; 122/1; 144/2; 210/1; 250/1  
C;Superfamily: CD27 antigen; NGF receptor repeat homology

Query Match 0.8%; Score 8; DB 2; Length 272;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623  
Db 123 CVPCPPGH 130  
|||||

RESULT 11  
T33466  
hypothetical protein C43H8.3 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T33466  
R;Tin-Wollam, A.M.  
submitted to the EMBL Data Library, October 1998  
A;Description: The sequence of C. elegans cosmid C43H8.  
A;Reference number: 221350  
A;Accession: T33466  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-281 <TIN>  
A;Cross-references: EMBL:AF098499; PIDN:AAC67398.1; GSPDB:GN00019; CESP:C43H8.3  
A;Experimental source: strain Bristol N2; clone C43H8  
C;Genetics:  
A;Gene: CESP:C43H8.3  
A;Map position: 1  
A;Introns: 58/1; 114/2; 242/3

Query Match 0.8%; Score 8; DB 2; Length 281;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 NITDFTVK 738  
Db 111 NITDFTVK 118  
|||||

RESULT 12  
A82929  
ATP synthase gamma chain UUI30 [imported] - Ureaplasma urealyticum  
C;Species: Ureaplasma urealyticum  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000

A;Accession: A82929  
R;Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.  
submitted to GenBank, February 2000  
A;Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mir  
A;Reference number: A82870  
A;Accession: A82929  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-286 <GLA>  
A;Cross-references: GB:AE002114; GB:AF222894; NID:g6899086; PIDN:AAF30536.1; GSPDB:GN001  
A;Experimental source: serovar 3; biovar 1  
C;Genetics:  
A;Gene: atpG; UUI30  
A;Genetic code: SGC3  
C;Superfamily: H+-transporting ATP synthase gamma chain

Query Match 0.8%; Score 8; DB 2; Length 286;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 EIVAGSDD 746  
|||||||  
Db 279 EIVAGSDD 286

RESULT 13  
S65471  
pyruvate decarboxylase (EC 4.1.1.1) (clone PDC2) - garden pea (fragment)  
C;Species: Pisum sativum (garden pea)  
C;Date: 19-Mar-1997 #sequence\_revision 06-Jun-1997 #text\_change 20-Jun-2000  
C;Accession: S65471; S65424  
R;Muecke, U.; Wohlfarth, T.; Fiedler, U.; Baeumlein, H.; Ruecknagel, K.P.; Koenig, S.  
Eur. J. Biochem. 237, 373-382, 1996  
A;Title: Pyruvate decarboxylase from Pisum sativum. Properties, nucleotide and amino acid  
A;Reference number: S65423; MUID:96215432; PMID:8647075  
A;Accession: S65471  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-405 <MUE>  
A;Cross-references: EMBL:266544; NID:g1177604; PIDN:CAA91445.1; PID:g1177605  
A;Accession: S65424  
A;Molecule type: protein  
A;Residues: 1-8,146-153;209-218;350-362;386-393 <MUW>  
C;Genetics:  
A;Gene: pdc  
C;Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain  
C;Keywords: carbon-carbon lyase; carboxy-lyase  
F;1-405/Product: pyruvate decarboxylase #status experimental <MAT>  
F;271-317/Domain: thiamin pyrophosphate-binding domain homology <TPB>

Query Match 0.8%; Score 8; DB 2; Length 405;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLY 276  
|||||||  
Db 50 KAVKPVLY 57

RESULT 14  
S57820  
pyruvate decarboxylase (EC 4.1.1.1) 1 - common tobacco (fragment)  
C;Species: Nicotiana tabacum (common tobacco)  
C;Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 15-Oct-1999  
C;Accession: S57820  
R;Bucher, M.; Brander, K.A.; Sbicego, S.; Mandel, T.; Kuhlemeier, C.  
Plant Mol. Biol. 28, 739-750, 1995  
A;Title: Aerobic fermentation in tobacco pollen.  
A;Reference number: S57819; MUID:95375236; PMID:7647304  
A;Accession: S57820  
A;Status: preliminary; nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-417 <BUC>

A;Cross-references: EMBL:X81854  
C;Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain  
C;Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 0.8%; Score 8; DB 2; Length 417;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLY 276  
|||||||  
Db 155 KAVKPVLY 162

RESULT 15  
AF0241  
probable coenzyme synthetase YPO1981 [imported] - Yersinia pestis (strain CO92)  
C;Species: Yersinia pestis  
C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
C;Accession: AF0241  
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.  
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel  
Nature 413, 523-527, 2001  
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A;Reference number: AB0001; MUID:21470413; PMID:11586360  
A;Accession: AF0241  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-428 <KUR>  
A;Cross-references: GB:AL590842; PIDN:CAC90794.1; PID:g15979994; GSPDB:GN00175  
C;Genetics:  
A;Gene: YPO1981

Query Match 0.8%; Score 8; DB 2; Length 428;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 604 ACALGSEQ 611  
|||||||  
Db 84 ACALGSEQ 91

Search completed: May 12, 2003, 13:41:38  
Job time : 19.4533 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:37:44 ; Search time 16.5146 Seconds  
(without alignments)  
1829.738 Million cell updates/sec

Title: US-10-073-333A-4  
Perfect score: 1027  
Sequence: 1 MLFRARGPVGRGWRPAEA.....KXKEDHFESVQLKTSRSPNI 1027

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	0.9	1487	2	US-08-760-489-2 Sequence 2, Appli
2	9	0.9	1487	2	US-08-760-489-4 Sequence 4, Appli
3	9	0.9	1487	4	US-09-185-373-2 Sequence 2, Appli
4	9	0.9	1487	4	US-09-185-373-4 Sequence 4, Appli
5	8	0.8	100	1	US-08-158-189-7 Sequence 7, Appli
6	8	0.8	133	4	US-08-468-560C-8 Sequence 8, Appli
7	8	0.8	139	2	US-08-219-237B-8 Sequence 8, Appli
8	8	0.8	140	4	US-08-477-347-17 Sequence 17, Appli
9	8	0.8	140	4	US-08-476-862-8 Sequence 8, Appli
10	8	0.8	153	4	US-09-286-529-2 Sequence 2, Appli
11	8	0.8	205	3	US-08-974-022-51 Sequence 51, Appli
12	8	0.8	205	4	US-08-795-445A-51 Sequence 51, Appli
13	8	0.8	205	4	US-08-795-447A-51 Sequence 51, Appli
14	8	0.8	205	4	US-08-974-186-51 Sequence 51, Appli
15	8	0.8	205	4	US-08-795-446B-51 Sequence 51, Appli
16	8	0.8	205	4	US-08-706-945D-138 Sequence 138, App
17	8	0.8	206	1	US-08-097-827-7 Sequence 7, Appli
18	8	0.8	206	1	US-08-494-574-7 Sequence 7, Appli
19	8	0.8	211	4	US-09-286-529-20 Sequence 20, Appli
20	8	0.8	299	4	US-09-286-529-17 Sequence 17, Appli
21	8	0.8	300	2	US-08-794-796-2 Sequence 2, Appli
22	8	0.8	438	1	US-08-097-827-11 Sequence 11, Appli
23	8	0.8	438	1	US-08-494-574-11 Sequence 11, Appli
24	7	0.7	15	4	US-08-602-999A-363 Sequence 363, App
25	7	0.7	15	4	US-09-500-124-363 Sequence 363, App
26	7	0.7	77	4	US-09-146-950-25 Sequence 25, Appli
27	7	0.7	100	2	US-08-467-822-23 Sequence 23, Appli

28	7	0.7	100	4	US-08-432-697-23	Sequence 23, Appli
29	7	0.7	100	4	US-08-466-248-23	Sequence 23, Appli
30	7	0.7	155	4	US-09-146-950-4	Sequence 4, Appli
31	7	0.7	159	4	US-09-146-950-20	Sequence 20, Appli
32	7	0.7	193	4	US-09-146-950-2	Sequence 2, Appli
33	7	0.7	197	4	US-09-146-950-18	Sequence 18, Appli
34	7	0.7	278	2	US-08-569-168-2	Sequence 2, Appli
35	7	0.7	283	4	US-08-509-024-2	Sequence 2, Appli
36	7	0.7	283	4	US-09-333-279-2	Sequence 2, Appli
37	7	0.7	283	4	US-09-072-993C-2	Sequence 2, Appli
38	7	0.7	283	5	PCT-US96-12374-2	Sequence 2, Appli
39	7	0.7	320	2	US-08-933-750C-12	Sequence 12, Appli
40	7	0.7	320	4	US-09-234-613-12	Sequence 12, Appli
41	7	0.7	331	4	US-09-086-483A-3	Sequence 3, Appli
42	7	0.7	383	2	US-08-569-168-7	Sequence 7, Appli
43	7	0.7	419	4	US-08-509-024-7	Sequence 7, Appli
44	7	0.7	419	4	US-09-333-279-7	Sequence 7, Appli
45	7	0.7	479	4	US-09-177-349-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1  
US-08-760-489-2  
; Sequence 2, Application US/08760489  
; Patent No. 5830696  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay M.  
; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC  
; TITLE OF INVENTION: ENZYMES  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/760,489  
; FILING DATE: 05-DEC-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/008,311  
; FILING DATE: 07-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haile, Ph.D., Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 09010/008001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5099  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1487 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; US-08-760-489-2

Query Match 0.9%; Score 9; DB 2; Length 1487;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 19 EAPRRGRSP 27  
|||||

Db 171 EAPRRGRSP 179

RESULT 2

US-08-760-489-4

; Sequence 4, Application US/08760489

; Patent No. 5830696

; GENERAL INFORMATION:

; APPLICANT: Short, Jay M.

; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC

; ENZYMES

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson, P.C.

; STREET: 4225 Executive Square, Suite 1400

; CITY: La Jolla

; STATE: CA

; COUNTRY: US

; ZIP: 92037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows95

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/760,489

; FILING DATE: 05-DEC-1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/008,311

; FILING DATE: 07-DEC-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Haile, Ph.D., Lisa A.

; REGISTRATION NUMBER: 38,347

; REFERENCE/DOCKET NUMBER: 09010/008001

; TELEPHONE: 619/678-5070

; TELEFAX: 619/678-5099

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1487 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FRAGMENT TYPE: internal

US-08-760-489-4

Query Match 0.9%; Score 9; DB 2; Length 1487;

Best Local Similarity 100.0%; Pred. No. 3.2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 EAPRRGRSP 27

Db 171 EAPRRGRSP 179

RESULT 3

US-09-185-373-2

; Sequence 2, Application US/09185373

; Patent No. 6335179

; GENERAL INFORMATION:

; APPLICANT: Short, Jay M.

; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC

; ENZYMES

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson, P.C.

; STREET: 4225 Executive Square, Suite 1400

; CITY: La Jolla

; STATE: CA

; COUNTRY: US

; ZIP: 92037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows95

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/185,373

; FILING DATE: 03-No. 6335179-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/760,489

; FILING DATE: 05-DEC-1996

; APPLICATION NUMBER: 60/008,311

; FILING DATE: 07-DEC-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Haile, Ph.D., Lisa A.

; REGISTRATION NUMBER: 38,347

; REFERENCE/DOCKET NUMBER: 09010/008001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 619/678-5070

; TELEFAX: 619/678-5099

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1487 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FRAGMENT TYPE: internal

; SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-185-373-2

Query Match 0.9%; Score 9; DB 4; Length 1487;

Best Local Similarity 100.0%; Pred. No. 3.2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 EAPRRGRSP 27

Db 171 EAPRRGRSP 179

RESULT 4

US-09-185-373-4

; Sequence 4, Application US/09185373

; Patent No. 6335179

; GENERAL INFORMATION:

; APPLICANT: Short, Jay M.

; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC

; ENZYMES

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson, P.C.

; STREET: 4225 Executive Square, Suite 1400

; CITY: La Jolla

; STATE: CA

; COUNTRY: US

; ZIP: 92037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows95

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/185,373

; FILING DATE: 03-No. 6335179-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/760,489

; FILING DATE: 05-DEC-1996

; APPLICATION NUMBER: 60/008,311

; FILING DATE: 07-DEC-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Haile, Ph.D., Lisa A.

; REGISTRATION NUMBER: 38,347

; REFERENCE/DOCKET NUMBER: 09010/008001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619/678-5070

TELEFAX: 619/678-5099

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1487 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-185-373-4

Query Match 0.9%; Score 9; DB 4; Length 1487;

Best Local Similarity 100.0%; Pred. No. 3.2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 EAPRRGRSP 27

|||||

Db 171 EAPRRGRSP 179

RESULT 5

US-08-158-189-7

Sequence 7, Application US/08158189

Patent No. 5641497

GENERAL INFORMATION:

APPLICANT: Bevins, Charles L.

APPLICANT: Jones, Douglas E.

TITLE OF INVENTION: Gastrointestinal Defensin Peptides,

TITLE OF INVENTION: CDNA Sequences, Methods for Production and Use Thereof

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris

STREET: One Liberty Place - 46th Floor

CITY: Philadelphia

STATE: PA

COUNTRY: U.S.A.

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/158,189

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/888,232

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Johnson, Philip S.

REGISTRATION NUMBER: 27,200

REFERENCE/DOCKET NUMBER: CH-0219

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-568-3100

TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 100 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-158-189-7

Query Match

Best Local Similarity 0.8%; Score 8; DB 1; Length 100;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 TAVLLVAL 943

|||||

Db 8 TAVLLVAL 15

RESULT 6

US-08-468-560C-8

Sequence 8, Application US/08468560C

Patent No. 6270998

GENERAL INFORMATION:

APPLICANT: NAGATA, Shigekazu

APPLICANT: ITOH, Naoto

APPLICANT: YONEHARA, Shin

TITLE OF INVENTION: DNA CODING FOR HUMAN CELL SURFACE

TITLE OF INVENTION: ANTIGEN

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP.

STREET: P.O. BOX 747

CITY: FALLS CHURCH

STATE: VA

COUNTRY: USA

ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/468,560C

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: MURPHY JR., GERLAD M.

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 20-4393P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-205-8000

TELEFAX: 703-205-8050

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 133 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-468-560C-8

Query Match

Best Local Similarity 0.8%; Score 8; DB 4; Length 133;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCCPPGH 623

|||||

Db 91 CVPCCPPGH 98

RESULT 7

US-08-219-237B-8

Sequence 8, Application US/08219237B

Patent No. 5874546

GENERAL INFORMATION:

APPLICANT: NAGATA, Shigekazu

APPLICANT: ITOH, Naoto

APPLICANT: YONEHARA, Shin

TITLE OF INVENTION: DNA Coding for Human Cell Surface Antigen

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: James W. Hellwege

STREET: P.O. Box 2266 Eads Station

CITY: Arlington

STATE: Virginia

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/219,237B  
FILING DATE: 28-MAR-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/872,129  
FILING DATE: 22-APR-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: James W. Hellwege  
REGISTRATION NUMBER: 28,808  
REFERENCE/DOCKET NUMBER: 516762  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 139 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-219-237B-8

Query Match 0.8%; Score 8; DB 2; Length 139;  
Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCCPGH 623  
|||||||  
Db 97 CVPCCPGH 104

RESULT 8

US-08-477-347-17  
Sequence 17, Application US/08477347  
Patent No. 6232446  
GENERAL INFORMATION:  
APPLICANT: WALLACH, David  
APPLICANT: BIGDA, Jacek  
APPLICANT: BELETSKY, Igor  
APPLICANT: METT, Igor  
TITLE OF INVENTION: TNF LIGANDS  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,347  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/115,685  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 106271  
FILING DATE: 08-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Townsend, G. Kevin  
REGISTRATION NUMBER: 34,033  
REFERENCE/DOCKET NUMBER: WALLACH-10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633

INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 140 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-477-347-17

Query Match 0.8%; Score 8; DB 4; Length 140;  
Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCCPGH 623  
|||||||  
Db 98 CVPCCPGH 105

RESULT 9

US-08-476-862-8  
Sequence 8, Application US/08476862  
Patent No. 6262239  
GENERAL INFORMATION:  
APPLICANT: WALLACH, David  
APPLICANT: BIGDA, Jacek  
APPLICANT: BELETSKY, Igor  
APPLICANT: METT, Igor  
APPLICANT: ENGELMANN, Hartmut  
TITLE OF INVENTION: TNF INHIBITORS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/476,862  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 107267  
FILING DATE: 12-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 94039  
FILING DATE: 06-APR-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 91229  
FILING DATE: 06-AUG-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 90339  
FILING DATE: 18-MAY-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: WALLACH-12A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 140 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-476-862-8



Query Match 0.8%; Score 8; DB 4; Length 140;  
Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623  
Db 98 CVPCPPGH 105

RESULT 10  
US-09-286-529-2  
; Sequence 2, Application US/09286529  
; Patent No. 6297367  
; GENERAL INFORMATION:  
; APPLICANT: Catherine Tribouley  
; TITLE OF INVENTION: NEW MEMBERS OF TNF AND TNFR FAMILIES  
; FILE REFERENCE: 1408.003/200130.439C1  
; CURRENT APPLICATION NUMBER: US/09/286,529  
; CURRENT FILING DATE: 1999-04-05  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 153  
; TYPE: PRT  
; ORGANISM: human  
US-09-286-529-2

Query Match 0.8%; Score 8; DB 4; Length 153;  
Best Local Similarity 100.0%; Pred. No. 4.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 PCPPGTFS 417  
Db 68 PCPPGTFS 75

RESULT 11  
US-08-974-022-51  
; Sequence 51, Application US/08974022  
; Patent No. 6015938  
; GENERAL INFORMATION:  
; APPLICANT: Boyle, William J.  
; APPLICANT: Lacey, David L.  
; APPLICANT: Calzone, Frank J.  
; APPLICANT: Chang, Ming-Shi  
; TITLE OF INVENTION: OSTEOPROTEGERIN  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: 1840 Dehavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/974,022  
; FILING DATE: 12-DEC-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/577,788  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.  
; REFERENCE/DOCKET NUMBER: A-378  
; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 205 amino acids

; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-974-022-51

Query Match 0.8%; Score 8; DB 3; Length 205;  
Best Local Similarity 100.0%; Pred. No. 5.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623  
Db 122 CVPCPPGH 129

RESULT 12  
US-08-795-445A-51  
; Sequence 51, Application US/08795445A  
; Patent No. 6284485  
; GENERAL INFORMATION:  
; APPLICANT: Boyle, William J.  
; APPLICANT: Lacey, David L.  
; APPLICANT: Calzone, Frank J.  
; APPLICANT: Chang, Ming-Shi  
; TITLE OF INVENTION: OSTEOPROTEGERIN  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: 1840 Dehavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/795,445A  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/577,788  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.  
; REFERENCE/DOCKET NUMBER: A-378  
; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 205 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-795-445A-51

Query Match 0.8%; Score 8; DB 4; Length 205;  
Best Local Similarity 100.0%; Pred. No. 5.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623  
Db 122 CVPCPPGH 129

RESULT 13  
US-08-795-447A-51  
; Sequence 51, Application US/08795447A  
; Patent No. 6284728  
; GENERAL INFORMATION:  
; APPLICANT: Boyle, William J.  
; APPLICANT: Lacey, David L.

APPLICANT: Calzone, Frank J.  
APPLICANT: Chang, Ming-Shi  
TITLE OF INVENTION: Osteoprotegerin  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: One Amgen Center Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91362-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/795,447A  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-378D2  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 205 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-795-447A-51  
Query Match 0.8%; Score 8; DB 4; Length 205;  
Best Local Similarity 100.0%; Pred. No. 5.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 616 CVPCCPGH 623  
Db 122 CVPCCPGH 129  
RESULT 14  
US-08-974-186-51  
Sequence 51, Application US/08974186  
Patent No. 6284740  
GENERAL INFORMATION:  
APPLICANT: Boyle, Willaim J.  
APPLICANT: Lacey, David L.  
APPLICANT: Calzone, Frank J.  
APPLICANT: Chang, Ming-Shi  
TITLE OF INVENTION: OSTEOPROTEGERIN  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: 1840 Dehavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,186  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/577,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.

REFERENCE/DOCKET NUMBER: A-378  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 205 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-974-186-51  
Query Match 0.8%; Score 8; DB 4; Length 205;  
Best Local Similarity 100.0%; Pred. No. 5.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 616 CVPCCPGH 623  
Db 122 CVPCCPGH 129  
RESULT 15  
US-08-795-446B-51  
Sequence 51, Application US/08795446B  
Patent No. 6288032  
GENERAL INFORMATION:  
APPLICANT: Boyle, Willaim J.  
APPLICANT: Lacey, David L.  
APPLICANT: Calzone, Frank J.  
APPLICANT: Chang, Ming-Shi  
TITLE OF INVENTION: OSTEOPROTEGERIN  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.  
STREET: 1840 Dehavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/795,446B  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/577,788  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Winter, Robert B.  
REFERENCE/DOCKET NUMBER: A-378  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 205 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-795-446B-51  
Query Match 0.8%; Score 8; DB 4; Length 205;  
Best Local Similarity 100.0%; Pred. No. 5.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 616 CVPCCPGH 623  
Db 122 CVPCCPGH 129  
Search completed: May 12, 2003, 13:42:24  
Job time : 18.5146 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 00:01:37 ; Search time 8811.67 Seconds  
(without alignments)  
11744.602 Million cell updates/sec

Title: US-10-073-333A-3  
Perfect score: 3556  
Sequence: 1 atgctgttcgcgcgcgcggg.....ggtaaaaaaaaaaaaaaaa 3556

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries.

Database : GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2959.2	83.2	3185	9	AK055902	AK055902 Homo sapi
2	1499.6	42.2	1737	6	AX060311	AX060311 Sequence
3	1182.6	33.3	1508	6	AX060301	AX060301 Sequence
4	991.4	27.9	3280	6	AX036667	AX036667 Sequence
5	989.8	27.8	3331	6	AX127726	AX127726 Sequence
6	987.8	27.8	3501	6	AX463904	AX463904 Sequence
7	860.8	24.2	2610	6	AX127742	AX127742 Sequence
8	860.8	24.2	2733	6	AX127740	AX127740 Sequence
9	732.8	20.6	2895	9	AK057647	AK057647 Homo sapi
10	711.6	20.0	2188	10	BC029010	BC029010 Mus muscu
11	687	19.3	2449	9	BC031648	BC031648 Homo sapi
12	633.8	17.8	2129	9	AK026832	AK026832 Homo sapi
13	594	16.7	5567	9	AB037745	AB037745 Homo sapi
14	518.4	14.6	136150	9	AC002081	AC002081 Homo sapi
15	504	14.2	576	6	AX336856	AX336856 Sequence
16	494	13.9	1129	6	AR184096	AR184096 Sequence
17	487	13.7	1717	6	AX013071	AX013071 Sequence
18	403	11.3	404	6	AX071721	AX071721 Sequence
19	381.8	10.7	1149	6	AX127746	AX127746 Sequence
20	294.2	8.3	1149	10	BC022655	BC022655 Mus muscu
21	285.8	8.0	1119	6	AX213279	AX213279 Sequence
22	285.8	8.0	1587	6	AX213277	AX213277 Sequence
23	265.4	7.5	101719	2	AL357495	AL357495 Homo sapi
24	234.6	6.6	627	6	AX127744	AX127744 Sequence
25	217	6.1	131321	2	AC080160	AC080160 Mus muscu
26	197.2	5.5	165865	2	AC079958	AC079958 Mus muscu
27	196.4	5.5	101719	2	AL357495	AL357495 Homo sapi
28	189	5.3	107889	9	AC004519	AC004519 Homo sapi
29	180.4	5.1	108063	2	AC116297	AC116297 Rattus no
30	169.2	4.8	677	6	AX036669	AX036669 Sequence
31	161.6	4.5	426	6	AX368203	AX368203 Sequence
32	140.6	4.0	437	6	AX182039	AX182039 Sequence
33	135.4	3.8	480	6	AX182004	AX182004 Sequence
34	130.4	3.7	29979	9	AC000122	AC000122 Homo sapi
35	127.2	3.6	131321	2	AC080160	AC080160 Mus muscu
36	126.4	3.6	466	6	AX261103	AX261103 Sequence
37	123.6	3.5	353	6	AX182022	AX182022 Sequence
38	118.6	3.3	166384	2	AC113446	AC113446 Mus muscu
39	118	3.3	108063	2	AC116297	AC116297 Rattus no
40	115.6	3.3	150129	2	AC119789	AC119789 Rattus no
41	115	3.2	399	6	AX182000	AX182000 Sequence
42	114	3.2	183815	2	AC099081	AC099081 Rattus no
43	104	2.9	443	6	AX127734	AX127734 Sequence
44	99.8	2.8	97114	9	AL356389	AL356389 Human DNA
45	94	2.6	265	6	AX368154	AX368154 Sequence

ALIGNMENTS

RESULT 1  
AK055902  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

AK055902 3185 bp mRNA linear PRI 01-AUG-2002  
Homo sapiens CDNA FLJ31340 fis, clone MESAN1000035, weakly similar  
to MAJOR SURFACE-LABELLED TROPHOZOITE ANTIGEN PRECURSOR.  
AK055902  
AK055902.1 GI:16550745  
oligo capping; fis (full insert sequence).  
Homo sapiens normal mesangial cells (NHMC56046-2) cDNA to mRNA,  
clone\_lib:MESAN1 clone:MESAN1000035.

Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,

Fukuzumi,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Irie,R.,  
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,  
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,  
Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,  
Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,  
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,  
Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.  
NEDO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 3185)  
Isogai,T., Otsuki,T. and Sugiyama,T.  
Direct Submission  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Submitted (24-Oct-2001) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: RAB and  
HRI.

FEATURES  
source  
1. .3185  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MESAN100035"  
/cell\_type="normal mesangial cells (NHMC56046-2)"  
/clone\_lib="MESAN1"  
/note="cloning vector: pME18SFL3-primary culture, normal  
mesangial cells"  
362. .1843  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB71041.1"  
/db\_xref="GI:16550746"  
/translation="MLEGNYQVLMKSGTNILYWRITGILMGSKAVKPVLVKNITIEGV  
AYTSECFPCPKPGTFSNKPQSFNCQVCPNRYSEKGAKECIRCKDDSOFSSEGSSECTE  
RPPCTTKDYFIHTPCDEEGKTQIMYKWIEPKICREDLTDAIRLPPSGEKKDCPPCNP  
GFYNNSSSCHPCPPGTESDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVG  
NSKCDGMNGWEVAGDHQSGAGGSDNDYLILNLHIPGFKPPTSMGTAGTSELGRITFV  
FETLCSADCVLFMVDINRKSTNVVESWGTEKEKQAYTHIFKNATFTFTWAFQRTNQ  
GQDNRFRINDMVKIYSITATNAVGVASSCRACALGSEQSGSSCVPCPPGHYIEKETN  
QCKECPDPTVLSIHQVYGKEACIPCGPGSKNNQDHSVCSYDCFFYPEKENQSLHYDFS  
NLSSVGSMLMNGPSFTSKGTYKTFHFNISYVGMGRRLWSVPTI"  
BASE COUNT 960 a 660 c 719 g 846 t  
ORIGIN

Query Match 83.2%; Score 2959.2; DB 9; Length 3185;  
Best Local Similarity 98.6%; Pred. No. 0;  
Matches 3032; Conservative 0; Mismatches 3; Indels 41; Gaps 3;  
QY 504 CAACCTCTCTTGGATCCCTCGTGGAAACTACATAGAATCTAATCGTGATGACTGCACGCT 563  
Db 111 CAGCTCTCTTGGATCCCTCGTGGAAACTACATAGAATCTAATCGTGATGACTGCACGCT 170  
QY 564 GTCTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGCTCTCTTTGAGTACCAGTA 623  
Db 171 GTCTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGCTCTCTTTGAGTACCAGTA 230  
QY 624 TGTCGACACAACATCTTCTTTGAGTCTCTTTATTCAAATGATCATAGTCCAGGAGATGGA 683  
Db 231 TGTCGACACAACATCTTCTTTGAGTCTCTTTATTCAAATGATCATAGTCCAGGAGATGGA 290  
QY 684 CACCACCACTGACAAGTGGGTAAACTTACAGACAATGGACAATGGGGCTCTCATTTCT-- 741  
Db 291 CACCACCACTGACAAGTGGGTAAACTTACAGACAATGGACAATGGGGCTCTCATTTCTGA 350  
QY 742 -----GTAATGCTGAAATCAGGCACAAACATAC 769  
Db 351 CAGCTCCTTAAATGCTTGAAGGCAATTACCAGGTAATGCTGAAATCAGGCACAAACATAC 410

QY 770 TCTACTGGAGAACTACAGGCATCCTTATGGGTCTTAAGCGGTCAAGCCTGTGCTGGTAA 829  
Db 411 TCTACTGGAGAACTACAGGCATCCTTATGGGTCTTAAGCGGTCAAGCCTGTGCTGGTAA 470  
QY 830 AAAATATCACAAATTGAAGGGGTGGCGTACACATCAGAAATGTCTTCCCTTGCAGGCCAGCA 889  
Db 471 AAAATATCACAAATTGAAGGGGTGGCGTACACATCAGAAATGTCTTCCCTTGCAGGCCAGCA 530  
QY 890 CATTTCAGCAACAAACCAGGTTCATTTCAACTGCCAGGTGTGCCAGAAACACCTATTCTG 949  
Db 531 CATTTCAGCAACAAACCAGGTTCATTTCAACTGCCAGGTGTGCCAGAAACACCTATTCTG 590  
QY 950 AGAAGGAGCCAAAGAATGTATAAGGTGTAAAGACGACTCTCAATTTTC-----AGGAT 1003  
Db 591 AGAAGGAGCCAAAGAATGTATAAGGTGTAAAGACGACTCTCAATTTTCAGAGGGAAGGAT 650  
QY 1004 CCAGTGAGTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCCATATC 1063  
Db 651 CCAGTGAGTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCCATATC 710  
QY 1064 CATGTGATGAAGAAAGAACACACAGATAATGTACAAGTGGATAGAGCCCCAAAATCTGCC 1123  
Db 711 CATGTGATGAAGAAAGAACACACAGATAATGTACAAGTGGATAGAGCCCCAAAATCTGCC 770  
QY 1124 GGAGGATCTCACAGATGCTATTAGATTGCCCTTCTGGAGAGAAAGGATTTGTCGCG 1183  
Db 771 GGAGGATCTCACAGATGCTATTAGATTGCCCTTCTGGAGAGAAAGGATTTGTCGCG 830  
QY 1184 CTTGCAACCTCGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCCTCGGAA 1243  
Db 831 CTTGCAACCTCGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCCTCGGAA 890  
QY 1244 CATTTCAGATGGAACCAAGAAATGTAGACCATGTCCAGAGGAACGGAGCCTGCACCTG 1303  
Db 891 CATTTCAGATGGAACCAAGAAATGTAGACCATGTCCAGAGGAACGGAGCCTGCACCTG 950  
QY 1304 GCTTTGAATATAAATGGTGGAAATGTCTTCTTGCCAAACATGAANAATCTTCTGCTCAATG 1363  
Db 951 GCTTTGAATATAAATGGTGGAAATGTCTTCTTGCCAAACATGAANAATCTTCTGCTCAATG 1010  
QY 1364 TTGGAATTCAAAGTCCGATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGA 1423  
Db 1011 TTGGAATTCAAAGTCCGATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGA 1070  
QY 1424 CTGGGCTGGAGGTTCTGACAAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTA 1483  
Db 1071 GTGGGCTGGAGGTTCTGACAAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTA 1130  
QY 1484 AACCAACACATCTATGACTGGAGCCACGGGTCTGAACTAGGAAGAATAACATTTGCT 1543  
Db 1131 AACCAACACATCTATGACTGGAGCCACGGGTCTGAACTAGGAAGAATAACATTTGCT 1190  
QY 1544 TTGAGACCTCTCTGTTTCAGCTGACTGTGTTTGTACTTTCATTTGATATTAATAGAAAA 1603  
Db 1191 TTGAGACCTCTCTGTTTCAGCTGACTGTGTTTGTACTTTCATTTGATATTAATAGAAAA 1250  
QY 1604 GTACAAATGTGGTAGAATCGTGGGTGGAAACCAAGAAAGAAACAAAGCTTACACCCATATCA 1663  
Db 1251 GTACAAATGTGGTAGAATCGTGGGTGGAAACCAAGAAAGAAACAAAGCTTACACCCATATCA 1310  
QY 1664 TCTTCAAGAATGCAACTTTTACATTTACATTTACATGGGCATTCAGAGAACTAATCAGGGTCAAG 1723  
Db 1311 TCTTCAAGAATGCAACTTTTACATTTACATTTACATGGGCATTCAGAGAACTAATCAGGGTCAAG 1370  
QY 1724 ATAATAGACGGTTTCATCAATGACATGGTGAAGATTTTATTCTATCACAGCCACTAATGCAG 1783  
Db 1371 ATAATAGACGGTTTCATCAATGACATGGTGAAGATTTTATTCTATCACAGCCACTAATGCAG 1430  
QY 1784 TTGATGGGGTGGCGTCTCTCATGCCGTGCCCTGTGCCCTCGGTTCTGAACAGTCCGGTTTCA 1843  
Db 1431 TTGATGGGGTGGCGTCTCTCATGCCGTGCCCTGTGCCCTCGGTTCTGAACAGTCCGGTTTCA 1490  
QY 1844 CGTGTGTCCCCTGCCCTCCAGGCCACTACATTGAGAAAGAAACCAACCAAGGAAAT 1903



Db 1491 CGTGTGTCCTCCCTCCAGGCCACTACATTGAGAAAAGAAACCAACCCAGTGAAGGAAT 1550  
QY 1904 GTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTCCAT 1963  
Db 1551 GTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTCCAT 1610  
QY 1964 GCGGGCCTGGGAGTAAACAAATCAGGACCAATTCGGTTTGCTATAGTGACTGCTTTTCT 2023  
Db 1611 GCGGGCCTGGGAGTAAACAAATCAGGACCAATTCGGTTTGCTATAGTGACTGCTTTTCT 1670  
QY 2024 ACCATGAAAAAGAAATCAGATTTTGCACATGACTTTAGCAACCTCAGCAGTGTGGCT 2083  
Db 1671 ACCCTGAAAAAGAAATCAGAGTTTGCACATGACTTTAGCAACCTCAGCAGTGTGGCT 1730  
QY 2084 CATTAATGAATGGCCCCAGCTTCACCTCCAAAGGAACAAATACCTCCATTTCAATA 2143  
Db 1731 CATTAATGAATGGCCCCAGCTTCACCTCCAAAGGAACAAATACCTCCATTTCAATA 1790  
QY 2144 TCAGTTTATGTGGCATGAGGGGAAGAAGATGGCTCTGTACCAACAAATATAACAGACT 2203  
Db 1791 TCAG-TTATGTGGCATGAGGGGAAGAAGATGGCTCTGTACCAACAAATATAACAGACT 1849  
QY 2204 TTACAGTAAAGAAATAGTGGCAGGGTCAGATGATTACACAAATTTGGTAGGGCATTTG 2263  
Db 1850 TTACAGTAAAGAAATAGTGGCAGGGTCAGATGATTACACAAATTTGGTAGGGCATTTG 1909  
QY 2264 TATGCCAGTCAACAATATTCTCTCTGAAAGTAAGGGTTTCCGAGCAGCCTTATCATCAC 2323  
Db 1910 TATGCCAGTCAACAATATTCTCTCTGAAAGTAAGGGTTTCCGAGCAGCCTTATCATCAC 1969  
QY 2324 AATCCATCATTTCTGCGAGATACATTTCCAGAGTACAGTTGAAACCAATGAAAAATA 2383  
Db 1970 AATCCATCATTTCTGCGAGATACATTTCCATAGGAGTCACAGTTGAAACCAATGAAAAATA 2029  
QY 2384 TTAATATAAAGAAAGATATGTTCCAGATTCCAAAGCCAAATACCAGATGTCATTTCT 2443  
Db 2030 TTAATATAAAGAAAGATATGTTCCAGATTCCAAAGCCAAATACCAGATGTCATTTCT 2089  
QY 2444 TTTATAAGTCTTCTACAGCAACAACATCTTGTTATTAATGGCCGATCAACTGTGAAAA 2503  
Db 2090 TTTATAAGTCTTCTACAGCAACAACATCTTGTTATTAATGGCCGATCAACTGTGAAAA 2149  
QY 2504 TGAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCAGCAAGTGCCCG 2563  
Db 2150 TGAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCAGCAAGTGCCCG 2209  
QY 2564 CAGGTACCTGTGATGGGTGTACGTTCTATTCTGTGGGAGAGTGCTGAAGCTTGCCCTC 2623  
Db 2210 CAGGTACCTGTGATGGGTGTACGTTCTATTCTGTGGGAGAGTGCTGAAGCTTGCCCTC 2269  
QY 2624 TGTGTACGGACATGACTTCCATGAGATTGAGGGAGCCTGCAAGAGAGGATTTCAAGAAA 2683  
Db 2270 TGTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCTGCAAGAGAGGATTTCAAGAAA 2329  
QY 2684 CCTTGATGTGTGGAATGAACCTAAATGGTGCAATTAAGGGAATTTCTTTGCCTGAGAAAA 2743  
Db 2330 CCTTGATGTGTGGAATGAACCTAAATGGTGCAATTAAGGGAATTTCTTTGCCTGAGAAAA 2389  
QY 2744 AGTTGGCAACTGTGAAACGGTTGACTTTTGGCTGAAGGTGGGAGCGGTGTGGGAGCTT 2803  
Db 2390 AGTTGGCAACTGTGAAACGGTTGACTTTTGGCTGAAGGTGGGAGCGGTGTGGGAGCTT 2449  
QY 2804 TTACTGCCGTTTGTCTGGTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATCAAAAACCTGG 2863  
Db 2450 TTACTGCCGTTTGTCTGGTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATCAAAAACCTGG 2509  
QY 2864 AATACAAATATTCCAAGTTAGTAATGACGACTAACTCAAAAGAGTGTTGAACCTCCCGGCTG 2923  
Db 2510 AATACAAATATTCCAAGTTAGTAATGACGACTAACTCAAAAGAGTGTTGAACCTCCCGGCTG 2569  
QY 2924 CAGACAGTTGTGCTATCATGGAAAGGAGAGATAATGAAGAGGAGTTGTATATTCCAATA 2983



Db 2570 CAGACAGTTGTGCTATCATGAAGGAGAAGATAATGAAGAGGAAGTTGTATATATCCAATA 2629  
QY 2984 AACAGTCACTACTAGGAAAAACTCAAATCTTTGGCAACCAAGGAAAAAGAACCATTTTG 3043  
Db 2630 AACAGTCACTACTAGGAAAAACTCAAATCTTTGGCAACCAAGGAAAAAGAACCATTTTG 2689  
QY 3044 AATCTGTTCAACTGAAAAACCTCAAGATCCCCCAATATATGAAGAGACAGTGTGTAGCCT 3103  
Db 2690 AATCTGTTCAACTGAAAAACCTCAAGATCCCCCAATATATGAAGAGACAGTGTGTAGCCT 2749  
QY 3104 TGAGACTAATGAACAAAGAAACCTGCTCTAGTTTACAGGACCATATTTTAGGGTCTGTC 3163  
Db 2750 TGAGACTAATGAACAAAGAAACCTGCTCTAGTTTACAGGACCATATTTTAGGGTCTGTC 2809  
QY 3164 CTCATACCTGTCAATGGTGATCTCACAGAGGAGGGCCCATGCCGCTGAAAAAGGAAGGA 3223  
Db 2810 CTCATACCTGTCAATGGTGATCTCACAGAGGAGGGCCCATGCCGCTGAAAAAGGAAGGA 2869  
QY 3224 GATTGAAACATTTGATTGCCCTTATCACATGGTCAAGTACCTTGCCAAATAAAGGAAAGCA 3283  
Db 2870 GATTGAAACATTTGATTGCCCTTATCACATGGTCAAGTACCTTGCCAAATAAAGGAAAGCA 2929  
QY 3284 AATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAAGAGATTTATCTGTATATA 3343  
Db 2930 AATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAAGAGATTTATCTGTATATA 2989  
QY 3344 CACATAACTGAAACCAAGTTTAAGCCCCACCACCAATGCACCTGCTGATGCATGCCATATAAT 3403  
Db 2990 CACATAACTGAAACCAAGTTTAAGCCCCACCACCAATGCACCTGCTGATGCATGCCATATAAT 3049  
QY 3404 AATGGGTAACTTTTATTCTTTATGATGTCTACATAACAAGTGTGATTTGGAAGGCACATG 3463  
Db 3050 AATGGGTAACTTTTATTCTTTATGATGTCTACATAACAAGTGTGATTTGGAAGGCACATG 3109  
QY 3464 TGAGCATATGCATATGATCCAATTTATGTTTCTTTTGTATATATTTGGGAAAAAT 3523  
Db 3110 TGAGCATATGCATATGATCCAATTTATGTTTCTTTTGTATATATTTGGGAAAAAT 3169  
QY 3524 AAAATTTTAAAGGT 3539  
Db 3170 AAAATTTTAAAGGT 3185

RESULT 2  
AX060311  
LOCUS AX060311 1737 bp DNA linear PAT 22-JAN-2001  
DEFINITION Sequence 19 from Patent WO0078802.  
ACCESSION AX060311  
VERSION AX060311.1 GI:12405800  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 1737)  
AUTHORS Shimkets,R.A., Fernandes,E., Vernet,C., Yang,M., Boldog,F.L. and Herrmann,J.L.  
TITLE Secreted polypeptides and corresponding polynucleotides  
JOURNAL Patent: WO 0078802-A 19 28-DEC-2000;  
Curagen Corporation (US)  
FEATURES  
source Location/Qualifiers  
1. .1737  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
296. .1690  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC24875.1"  
/db\_xref="GI:12405801"  
/translation="MKNQVCSKCGEGTYSLGSGIKFDEWDELPAQFSNIATFMDTVVG  
PSDSRPDGCNNSSWTPRGNIESNRDDCTVSLIYVHLKKSGYVFVEYQYVDNNIFFE  
FFIQNDQCQEMDITTDKWKVLTNDNGEWSHVMKSLGNTILYWRITGILMGSAVKPV  
LVKNITIEGVAYTSECFCKPGTFSNKPQSFNCQVCPNRTYSEKGAKECIRCKDDSQF

SEGSSECTERPPCTTKDYFQIHTPCDEEGKTOIMYKWKIEPKICREDLTDAILRPPSG  
EKKDCPPCNPGFYNNGSSSCHPCPPGTFSDGTEKCRPCPAGTEPALGFEYKWNVLPG  
NMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTGAT  
GSELGRITFVFETLCSADCVLVFMVDINRKSTNVVESWGGTKERQAYTHIIFKNATFT  
FTWGIPIRELIQGP"  
BASE COUNT 495 a 381 c 409 g 451 t 1 others  
ORIGIN

Query Match 42.2%; Score 1499.6; DB 6; Length 1737;  
Best Local Similarity 98.7%; Pred. No. 0;  
Matches 1545; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

QY	189	GAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC	248
Db	142	GAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC	201
QY	249	CATTCCAAATTCTGCAGTGGACTGCTCTGGCCTGCCTGACCCAGTGAGAGGCCAAAGAATG	308
Db	202	CATTCCAAATTCTGCAGTGGACTGCTCTGGCCTGCCTGACCCAGTGAGAGGCCAAAGAATG	261
QY	309	CACTTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGCAGTAAGTG	368
Db	262	CACTTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGCAGTAAGTG	321
QY	369	TGGTGAAGCACCTATTCCCTTGGGAGTGGCATCAAATTTGATGAATGGGATGAATTGCC	428
Db	322	TGGTGAAGCACCTATTCCCTTGGGAGTGGCATCAAATTTGATGAATGGGATGAATTGCC	381
QY	429	GGCAGGATTTCTAAACATCGCAACATTCATGGACACTGTGTGGGCCCTTCTTGACAGCAG	488
Db	382	GGCAGGATTTCTAAACATCGCAACATTCATGGACACTGTGTGGGCCCTTCTTGACAGCAG	441
QY	489	GCCAGACGGCTGTAACAACTCTTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGCTT	548
Db	442	GCCAGACGGCTGTAACAACTCTTTTGATCCCTCGTGGAACCTACATAGAATCTAATCG	501
QY	549	TGATGACTGCACGGTGTCTTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGCTT	608
Db	502	TGATGACTGCACGGTGTCTTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGCTT	561
QY	609	CTTTGAGTACCAGTATGTCGACAACAACATCTTCTTTGAGTTCITATTCAAATGATCA	668
Db	562	CTTTGAGTACCAGTATGTCGACAACAACATCTTCTTTGAGTTCITATTCAAATGATCA	621
QY	669	GTGCCAGGAGATGGACACCACCACCTGACAAGTGGGTAAACCTTACAGACAATGGAGAATG	728
Db	622	GTGCCAGGAGATGGACACCACCACCTGACAAGTGGGTAAACCTTACAGACAATGGAGAATG	681
QY	729	GGGCTCTCATCTGTAATGCTGAAATCAGGCACAAACATACTCTACTGGAGAACTACAGG	788
Db	682	GGGCTCTCATCTGTAATGCTGAAATCAGGCACAAACATACTCTACTGGAGAACTACAGG	741
QY	789	CATCCTTATGGGTTCTAAGCGCGTCAAGCCTGTCTGGTAAAAAATATCACAATTTGAAGG	848
Db	742	CATCCTTATGGGTTCTAAGCGCGTCAAGCCTGTCTGGTAAAAAATATCACAATTTGAAGG	801
QY	849	GGTGGCGTACACATCAGAAATGTTTTCCTTGCAGCCAGGCACATTCAGCAACAAACCAGG	908
Db	802	GGTGGCGTACACATCAGAAATGTTTTCCTTGCAGCCAGGCACATTCAGCAACAAACCAGG	861
QY	909	TTCAATTCAACTGCCAGGTGTGTCCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAATG	968
Db	862	TTCAATTCAACTGCCAGGTGTGTCCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAATG	921
QY	969	TATAAGGTGTAAAGACGACTCTCAATTTTTC-----AGGATCCAGTGAAGTGTACAGAGCG	1022
Db	922	TATAAGGTGTAAAGACGACTCTCAATTTTTCAGAGGAAGGATCCAGTGAAGTGTACAGAGCG	981
QY	1023	CCCTCCCTGTACCAAAAAGACTATTTCCAGATCCATACTCCATGTGTGAAGAAAGGAAA	1082
Db	982	CCCTCCCTGTACCAAAAAGACTATTTCCAGATCCATACTCCATGTGTGAAGAAAGGAAA	1041
QY	1083	GACACAGATAATGTACAAGTGGATAGAGCCCCAAAAATCTGCCGGGAGGATCTCACAGATGC	1142

Db	1042	GACACAGATAATGTACAAAGTGGATAGAGCCCCAAAATCTGCCGGGAGGATCTCACAGATGC	1101
QY	1143	TATTAGATTGCCCCCTCTCTGGAGAGAAAGGATTGTCCGCCCTTGCAACCCCTGGATTTTTA	1202
Db	1102	TATTAGATTGCCCCCTCTCTGGAGAGAAAGGATTGTCCGCCCTTGCAACCCCTGGATTTTTA	1161
QY	1203	TAACAAATGGATCATCTTCTGCCATCCCTGTCCCTCCCTGGAAACATTTTCAGATGGAAACAA	1262
Db	1162	TAACAAATGGATCATCTTCTTGCCATCCCTGTCCCTCCCTGGAAACATTTTCAGATGGAAACAA	1221
QY	1263	AGAAATGTAGACCATGTCCAGCAGGAACGGAGCCCTGCACITTTGGCTTTTGAATATAAAATGGTG	1322
Db	1222	AGAAATGTAGACCATGTCCAGCAGGAACGGAGCCCTGCACITTTGGCTTTTGAATATAAAATGGTG	1281
QY	1323	GAATGTCTCTTCCCTGGCAACATGAAAACTTCCCTGCTTCAATGTGGGAATTCAAAAGTGGCA	1382
Db	1282	GAATGTCTCTTCCCTGGCAACATGAAAACTTCCCTGCTTCAATGTGGGAATTCAAAAGTGGCA	1341
QY	1383	TGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGA	1442
Db	1342	TGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGA	1401
QY	1443	CAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCCCAACATCTATGAC	1502
Db	1402	CAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCCCAACATCTATGAC	1461
QY	1503	TGGAGCCACGGGTTCTGAACCTAGGAAGAATAACATTTTGTCTTTGAGACCCCTCTGTTTCAGC	1562
Db	1462	TGGAGCCACGGGTTCTGAACCTAGGAAGAATAACATTTTGTCTTTGAGACCCCTCTGTTTCAGC	1521
QY	1563	TGACTGTGTTTGTACTTTCATGGTGGATATTAAAGAAAAAGTACAAAATGTGGTAGAATC	1622
Db	1522	TGACTGTGTTTGTACTTTCATGGTGGATATTAAAGAAAAAGTACAAAATGTGGTAGAATC	1581
QY	1623	GTGGGTTGGAACCAAGAAAAACAAGCTTACACCCCATATCATCTTCAAGAANTGCAACCTTT	1682
Db	1582	GTGGGTTGGAACCAAGAAAAACAAGCTTACACCCCATATCATCTTCAAGAANTGCAACCTTT	1641
QY	1683	TACATTTACATGGG--CATTCAGAGAACTAAT--CAGGGTCAAGATAANTAGACGGTTCA	1738
Db	1642	TACATTTACATGGGGCATTTCCAGAGAACTAATTCAGGGTCCCAAGATAANTAGACGGTTCC	1701
QY	1739	TCAAT 1743	
Db	1702	NCCAT 1706	

RESULT 3  
AX060301  
LOCUS AX060301 1508 bp DNA linear PAT 22-JAN-2001  
DEFINITION Sequence 9 from Patent WO0078802.  
ACCESSION AX060301  
VERSION AX060301.1 GI:12405790  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1508)  
AUTHORS Shimkets,R.A., Fernandes,E., Vernet,C., Yang,M., Boldog,F.L. and Herrmann,J.L.  
TITLE Secreted polypeptides and corresponding polynucleotides  
JOURNAL Patent: WO 0078802-A 9 28-DEC-2000;  
FEATURES  
Location/Qualifiers  
source 1..1508  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
226..1461  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC24870.1"

/db\_xref="GI:12405791"  
/translation="MHSSWIPRGNYYIESNRDDCTVSLIYAVHLKKSgyvffeyqyvDN  
NIFEFFIQNDQCQEMDTTDDKWKLTNDNGEWSHSMVLKSGTNILYRTTGILMGSK  
AVKPVLVKNITIEGVAYTSECFPCPKPGTFESNKPFSFNCQVCPRNTYSEKAKECIRCK  
DQSOFSEESSECTERPPCTTKDYFOIHTPCDEEGTKQIMYKWIPIKICREDLTDAIR  
LPPSGEKKDCPPCNPGEFYNNGSSSCHPCPPGTFSDGTECRPCPAGTEPALGFEYKWW  
NVLPGNMKTSFCNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIFPGFKPPTS  
MTGATGSELGRITVFETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFK  
NATFTFTWGIPIRELIQPR"  
BASE COUNT 444 a 306 c 355 g 402 t 1 others  
ORIGIN

Query Match 33.3%; Score 1182.6; DB 6; Length 1508;  
Best Local Similarity 98.4%; Pred. No. 7.5e-267;  
Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

QY 506 ACTCTTCTTGATCCCTCGTGGAAACTACATAGAATCTAATCGTGATGACTGCACGGGTG 565  
Db 230 ACTCTTCTTGATCCCTCGTGGAAACTACATAGAATCTAATCGTGATGACTGCACGGGTG 289  
QY 566 CTTTGATCTATGCTGTGCACCTTAAGAAAGTCAGGCTATGCTCTCTTTGAGTACCAGTATG 625  
Db 290 CTTTGATCTATGCTGTGCACCTTAAGAAAGTCAGGCTATGCTCTCTTTGAGTACCAGTATG 349  
QY 626 TCGACAACAACATCTTCTTTGAGTCTTTTATTCAAAATGATCAGTGCCAGGAGATGGACA 685  
Db 350 TCGACAACAACATCTTCTTTGAGTCTTTTATTCAAAATGATCAGTGCCAGGAGATGGACA 409  
QY 686 CCACCACTGACAAAGTGGGTAAAACTTACAGACAATGGAGATGGGGCTCTCATTTCTGTAA 745  
Db 410 CCACCACTGACAAAGTGGGTAAAACTTACAGACAATGGAGATGGGGCTCTCATTTCTGTAA 469  
QY 746 TGCTGAATCAGGCACAAACATACTCTACTGGAGAACTACAGGCATCCTTATGGGTTCTA 805  
Db 470 TGCTGAATCAGGCACAAACATACTCTACTGGAGAACTACAGGCATCCTTATGGGTTCTA 529  
QY 806 AGGCGGTCAAGCCTGTGCTGTGTAAATAATATCACAAATGAAGGGTGGCGTACACATCAG 865  
Db 530 AGGCGGTCAAGCCTGTGCTGTGTAAATAATATCACAAATGAAGGGTGGCGTACACATCAG 589  
QY 866 AATGTTTCCCTTGCAAGCCAGGCACATTCAGCAACAAACCAGGTTTCAATCAACTGCCAGG 925  
Db 590 AATGTTTCCCTTGCAAGCCAGGCACATTCAGCAACAAACCAGGTTTCAATCAACTGCCAGG 649  
QY 926 TGTGTCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAAATGTATAAGGTGTAAAGACG 985  
Db 650 TGTGTCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAAATGTATAAGGTGTAAAGACG 709  
QY 986 ACTCTCAATTTTC-----AGGATCCAGTGAGTGTACAGAGCGCCCTCCCTGTACCACAA 1039  
Db 710 ACTCTCAATTTTCAGAGGAAGGATCCAGTGAGTGTACAGAGCGCCCTCCCTGTACCACAA 769  
QY 1040 AAGACTATTTCAGATCCATACTCCATGTGATGAAGAGAAAGACACAGATAATGTACA 1099  
Db 770 AAGACTATTTCAGATCCATACTCCATGTGATGAAGAGAAAGACACAGATAATGTACA 829  
QY 1100 AGTGGATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGATGCTATTAGATTGCCCCCTT 1159  
Db 830 AGTGGATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGATGCTATTAGATTGCCCCCTT 889  
QY 1160 CTGGAGAGAAGAGGATTGTCCGCCCTTGCAACCCCTGGATTTTATAACAATGGATCATCTT 1219  
Db 890 CTGGAGAGAAGAGGATTGTCCGCCCTTGCAACCCCTGGATTTTATAACAATGGATCATCTT 949  
QY 1220 CTTGCCATCCCTGTCTCTCGAAACATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1279  
Db 950 CTTGCCATCCCTGTCTCTCGAAACATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1009  
QY 1280 CAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAATGGTGAATGTCTCTTCCTGGCA 1339  
Db 1010 CAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAATGGTGAATGTCTCTTCCTGGCA 1069  
QY 1340 ACATGAAAACTTCCTGCTTCAATGTTGGGAATTCAAAGTCCGATGGAATGAATGGTTGGG 1399

Db 1070 ACATGAAAACTTCCTGCTTCAATGTTGGGAATTCAAAGTCCGATGGAATGTTGGG 1129  
QY 1400 AGTGCGTGGAGATCATATCCAGAGTGGGCTGGAGGTTCTTGACAATGATTACCTGATCT 1459  
Db 1130 AGTGCGTGGAGATCATATCCAGAGTGGGCTGGAGGTTCTTGACAATGATTACCTGATCT 1189  
QY 1460 TAAACTTGATATCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGTTCTG 1519  
Db 1190 TAAACTTGATATCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGTTCTG 1249  
QY 1520 AACTAGGAGAATAACATTTTGTCTTTTGAGACCCCTCTGTTTCAGCTGACTGTGTTGTACT 1579  
Db 1250 AACTAGGAGAATAACATTTTGTCTTTTGAGACCCCTCTGTTTCAGCTGACTGTGTTGTACT 1309  
QY 1580 TCATGGTGGATATTAAATAGAAAAAGTACAAATGTGGTAGAATCGTGGGTGGAAACCAAAG 1639  
Db 1310 TCATGGTGGATATTAAATAGAAAAAGTACAAATGTGGTAGAATCGTGGGTGGAAACCAAAG 1369  
QY 1640 AAAACAAGCTTACACCCATATCATCTTCAAGAAATGCAAACTTTTACATTTACATGGG--C 1697  
Db 1370 AAAACAAGCTTACACCCATATCATCTTCAAGAAATGCAAACTTTTACATTTACATGGGGCA 1429  
QY 1698 ATTCCAGAGAACTAAT--CAGGGTCAAGATAANTAGACGGTTTCATCAAT 1743  
Db 1430 TTCCAGAGAACTAATTCAGGGTCCAAAGATAATAGACGGTTCCNCCAT 1477

RESULT 4

AX036667

LOCUS AX036667 3280 bp DNA linear PAT 16-NOV-2000

DEFINITION Sequence 1 from Patent WO0058460.

ACCESSION AX036667

VERSION AX036667.1 GI:11226242

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 3280)

AUTHORS Bruck,C.E., Vinals,Y.D., Coche,T. and Cassart,J.P.

TITLE Novel compounds

JOURNAL Patent: WO.0058460-A 1 05-OCT-2000;

BRUCK CLAUDINE ELVIRE MARIE (BE) ; SMITHKLINE BEECHAM BIOLOG (BE) ;

VINALS Y DE BASSOLS CARLOTA (BE) ; COCHE THIERRY (BE) ; CASSART

JEAN POL (BE)

FEATURES Location/Qualifiers

Source 1..3280

/organism="Homo sapiens"

/db\_xref="taxon:9606"

BASE COUNT 810 a 930 c 797 g 743 t

ORIGIN

Query Match 27.9%; Score 991.4; DB 6; Length 3280;

Best Local Similarity 60.6%; Pred. No. 7.5e-222;

Matches 1758; Conservative 0; Mismatches 1116; Indels 27; Gaps 7;

QY 171 GCTTCCTCCTTGCCAGGAGAAAGATTATCATTGAAATATACGGAATGTGATAGCAGTGG 230

Db 135 GCTTCATGCCCTGCAAAGAGCTGTGAGTACCACCTATGAGTACACGGCGTGTGACAGCACGGG 194

QY 231 CTCCAGGTGGAGAGTTGCCATTCCAAATTTGCAAGTGGACTGCTCTGGCCCTGCCTGACCC 290

Db 195 TTCCAGGTGGAGGTCGCCGTGCCGCATACCCCGGCCCTGTGCACACAGCCTGCCTGACCC 254

QY 291 AGTGACAGGCAAGAAATGCACCTTTCTCTCTGTGCTTCTTGAGAGTATCTAGAAATGAAGAA 350

Db 255 CGTCAAGGGCACCGAGTGTCTCTCTCTGCAACGCCCGGGAGTTTCTGGATATGAAGGA 314

QY 351 CCAGGTATGCAGTAGTGTGGTCAAGGCACCTATTCCCTTGGGAGTGGCATCAAATTTGA 410

Db 315 CCAGTCATGTAAGCCATGCGCTGAGGGCCGCTACTCCCTCGGCACAGGCATTCGGTTTGA 374



QY 411 TGAATGGGATGAATTGCCGGCAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGT 470  
Db 375 TGAGTGGGATGAGCTGCCCATGGCTTTGCCAGGCTCTCAGCCAAACATGGAGCTGGATGA 434  
QY 471 GGGCCCTTCTGACAGAGGCCAGACGGCTGTAAACACTCTTCTTGGATCCCTCGTGGAAA 530  
Db 435 CAGTGTCTGAGTCCA--CCGGGAACGTGTACTTCGTCCAAGTGGGTTCCCGGGGCGA 491  
QY 531 CTACATAGAATCTAATCGTGATGACTGCACGGTGCTTTTGATCTATGCTGTGCACCTTAA 590  
Db 492 CTACATCGCTCCAACACGAGCAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 551  
QY 591 GAAGTCAGGCTATGTCTTTTGAGTACCAGTATGTCGACAACAACATCTTCTTTGAGTT 650  
Db 552 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 611  
QY 651 CTTTATTCAAAATGATGCCAGGAGATGGACACCACCACCTGACAAGTGGGTAAACT 710  
Db 612 TTTCTGTTCAGAAATGACCAAGTGCACGCCCAATGCAGATGACTC---CAGGTGGATGRAGAC 668  
QY 711 TACAGACAATGGAGAATGGGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATACT 770  
Db 669 CACAGAAA--AGGATGGGAATCCACAGCTGTGGAGCTAAATCGAGGCCAATAATGTCTT 725  
QY 771 CTACTGGAGAACTACAGGCAATCCTTATGGGTTCTAAGCGGTCAAGCCTGTGCTGTAAA 830  
Db 726 CTATTGGAGAACACACAGCCTTCTCAGTATGGACCAAGTACCCCAAGCCTGTGCTGTGAG 785  
QY 831 AAATATCACAAATGAAGGGGTGGGTGACATCAGAAATGTTTCTTGCAGGCCAGGCAC 890  
Db 786 AAACATTGCCATAACAGGGGTGGCTTACACTTCAGAATGCTTCCCTGCAACCTGGCAC 845  
QY 891 ATTCAGCAACAAACAGGTTTCAATCAACTGCCAGGTGTGCCAGAAACACCTATTCTGA 950  
Db 846 GTATGACAGACAAGCAGGGCTCCTTTCTGCAAACTTTGCCAGCCTTCTTATTCAAA 905  
QY 951 GAAAGGAGCCAAAGAAATGTATAAGTGTAA--AGACGACTCTCAATTTTCAGGATCCAG 1007  
Db 906 TAAAGGAGAAACTTCTTGCCACCAGTGTGACCCTGCACAAATCTCAGAGAAAGGATCTTC 965  
QY 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAGACTATTTCAGATCCATCTCCATG 1067  
Db 966 TTCTCTGTAACGTGCGCCGAGCTTGACAGACAAAGATTATTTCACACACACACGGCCTG 1025  
QY 1068 TGATGAAGAAGGAAAGACACAGATAATGTACAAGTGGATAGACCCCAAAATCTGCCGGGA 1127  
Db 1026 CGATGCCAACGGAGAGACACAACACTCATGTACAAATGGGCCAAGCCGAAATCTGTAGCGA 1085  
QY 1128 GGATCTCAGAGTGTATTAGATTGCCCTTCTTGCCATCCCTGTCTCTCTGGAACATT 1247  
Db 1086 GGACCTTGAGGGGCGAGTGAAGCTGCCTCTGCTGTGAGAGACCCACTGCCACCCCTG 1145  
QY 1188 CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTCTGGAACATT 1247  
Db 1146 CAACCCAGGCTTCTTCAAAACCAACAACAGCACCTGCCAGCCCTGCCATATGTTCCCTA 1205  
QY 1248 TTCAGATGGAAACCAAGAAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTGGCTT 1307  
Db 1206 CTCCAATGG---CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGATT 1262  
QY 1308 TGAATATAATGGTGGAAATGTCTTCTTGGCAACATGAAAACCTTCTCTGCTTCAATGTTGG 1367  
Db 1263 TGAATACAAATGGTGGAAACACGCTGCCCAACAACATGGAAACGACCGCTTCTCAGTGGGAT 1322  
QY 1368 GAATTCAAAGTCGATGGAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427  
Db 1323 CAACCTTCGAGTACAAGGGCATGACAGGCTGGAGGTGGCTGGTGATACATTTACACAGC 1382  
QY 1428 GGCTGGAGGTTCTGACAAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAACC 1487  
Db 1383 TGCTGGAGCCTCAGACAATGACTTTCATGATCTCACTCTGTTGTCAGGATTTAGACC 1442  
QY 1488 ACCAACATCTATGACTGGAGCCAC---GGGTTCTGAACCTAGGAAGAATAACATTTGTCTT 1544

Db 1443 TCCGCAGTCGGTGATGGCAGACACAGAGAAATAAGAGGGTGGCCAGAATCACATTTGTCTT 1502  
QY 1545 TGAGACCCCTCTGTTTCAGCTGACTGTGTTTTTGTACTTCATGGTGGATATTAAATAGAAAAG 1604  
Db 1503 TGAGACCCCTCTGTTCTGTGAACCTGTGAGCTCTACTTCATGGTGGGTGTGAATCTAGGAC 1562  
QY 1605 TACAAATGTGGTAGAATCGTGGGGTGGAAACCAGAAAGAAACAAAGCTTACACCCATATCAT 1664  
Db 1563 CAACACTCTGTGGAGACGTGGAAAAGTTCCAAAGGCAACACAGTCCTATACCTACATCAT 1622  
QY 1665 CTTCAAGAAATGCAACTTTTACATTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGA 1724  
Db 1623 TGAGGAGAACACTACCACGAGCTTCACCTGGGCCTTCCAGAGGACCACCTTTTCATGAGGC 1682  
QY 1725 TAATAGACGGTTCATCAATGACATGGTGAAGATTTTATCTATCAGAGCCACATAATGACAT 1784  
Db 1683 AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCAACCAATGTTAT 1742  
QY 1785 TGATGGGGTGGCGTCCCTCATTCGCGTGCCCTGTGCCCTTCTGAACAGTTCGGGTTCATC 1844  
Db 1743 GAATGGCGTGGCCTCCTACTGCGCTCCCTGTGCCCTAGAAAGCCTCTGATGTGGGCTCCTC 1802  
QY 1845 GTGTGTCCCTGCCCTCCAGGCCACTACATTCAGGTCTATGGCAAGAGGCTTGTATTCATG 1904  
Db 1803 CTGCACCTCTTGTCTGTCTGTACTATATGTACCGAGATTGACGAACTGCCACTCCTG 1862  
QY 1905 TCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTCATG 1964  
Db 1863 CCCCCCTAACACAATCTGAAAGCCCAACAGCCTTATGGTGTCCAGGCTGTGTGCCCTG 1922  
QY 1965 CGGCGCTGGGAGTAAAAACAATCAGGACCATTCCGTTTGTATAGTGAATGCTTTCCTA 2024  
Db 1923 TGGTCCAGGGACCAAGAACAACAAGATCCACTCTCTGTGTCTACAAATGATTCACCTTCTC 1982  
QY 2025 CCAATGAAAAGAAAATCAGATTTTGCACATATGACTTTAGCAACCTCAGCAGTGTGGGCTC 2084  
Db 1983 ACGCAACACTCCAACCCAGGACTTTCAACTACAACCTTCTCCGCTTTGGCAACACACCGTCA 2042  
QY 2085 ATTAATGAATGGCCCCAGCTTCACCTCCAAGGAACAATAACTTCCATTTCTTCAATAT 2144  
Db 2043 TCTTGTCTGGAGGCCAAGCTTCACTTCCAAGGGTTGAAATACTTCCATCACCTTACCCCT 2102  
QY 2145 CAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTT 2204  
Db 2103 CAGTCTCTGTGGAAACCAGGGTAGGAAAATGTCTGTGTGCACCCGACAAATGTCACTGACCT 2162  
QY 2205 TACAGTAAAAGAAAATAGTGGCAGGGTFCAGATGATTAACAAAATTTGGTAGGGGCAATTTGT 2264  
Db 2163 CCGGATTCTCTGAGGGTGAGTCAGGG-----TTCTCCAAATCTATCACAGCCTTACGT 2213  
QY 2265 ATGCCAGTCAACAATTTATTCCTTCTGAAAGTAAGGGTTTCCGAGCAGCCTTATCATCACA 2324  
Db 2214 CTGCCAGGCACTCATCATCCCCCAGAGGTGACAGGCTACAAGSCGGGGTTTCCCTCACA 2273  
QY 2325 ATCCATCATCTTGGCAGATACATTCATAGGAGTCACAGTTGAAACCACATTTGAAAAATAT 2384  
Db 2274 GCCTGTGAGCCTTGTGTGATCGACTTATTTGGGTGACAACAGATATGACTCTGGATGGAAT 2333  
QY 2385 TAATATAAAGAAGATATGTTCCAGTTTCCAACAAGCCAAATACCAGATGTGCATTTCTT 2444  
Db 2334 CACCTCCCCAGCTGAACCTTTTCCACCTGGAGTCCCTTGGGAATACCGGACGTGATCTTCTT 2393  
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAATGGCCGATCAACTGCTGTAATAAT 2504  
Db 2394 TTATAGGTCCAATGATGTGACCCAGTCTCTCAGTTCTTGGAGATCAACACCATCCCGCT 2453  
QY 2505 GAGGTGTAATCCTACTAAAATCTGGAGCAGAGATGATTTTCACTCCCCAGCAAGTGCCACG 2564  
Db 2454 CAGGTGCAGTCCACAGAAAACCTGTCCCTGGAAGTTTGTGCTGCCAGGAACGTGCTCAGA 2513  
QY 2565 AGGTACCTGTGATGGGTGTACGTTCTATTTCTCTGTGGGAGAGTGTGAAGCTTGCCCTCT 2624



Db 2514 TGGGACCTGTGATGGCTGCAACTTCCACTTCCTGTGTGGAGAGCGCGGCTGTGCCCCGT 2573

QY 2625 GTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCCTGCAAGAGAGAGGATTTCAGGAAAC 2684

Db 2574 CTGCTCAGTGGCTGACTACCATGCTATCGTCAGCAGCTGTGTGGCTGGATCCAGAAGAC 2633

QY 2685 CTTGTATGTGTGGAATGAACCTAAATGGTGCATTAAAGGAATTTCTTGCCTGAGAAAA 2744

Db 2634 TACTTACGTGTGGCGAGAACCCCAAGCATGCTGTGGTCATTTCTCTGCTGAGCAGAG 2693

QY 2745 GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGGTGGGAGCCGCTGTGGGAGCTTT 2804

Db 2694 AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG 2753

QY 2805 TACTGCCGTTTGGCTGGTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATCAAAAACCTGGA 2864

Db 2754 TACTGCCATCTGCTCACCGTCTTGACCTGCTACTTTTGGAAAAAGAAATCAAAAACCTAGA 2813

QY 2865 ATACAAATATTCCAAAGTTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCCGGCTGC 2924

Db 2814 GTACAAGTACTCCAAGCTGGTGATGAATGCTACTCTCAAGGACTGTGACCTGCCAGCAGC 2873

QY 2925 AGACAGTTGTGCTATCATGGAAGGAGAGAGATAATGAAGAGGAAGTTGTATATTTCCAATAA 2984

Db 2874 TGACAGCTGCGCCATCATGGAAGGCGAGGATGTAGAGGACGACCTCATCTTTACCAGCAA 2933

QY 2985 ACAGTCACTACTAGGAAAACTCAAACTTTTGGCAACCAAGGAAAAAGAGACCATTTTGA 3044

Db 2934 GAAGTCACTCTTTGGGAAGATCAAAATCATTTACCTCCAAGAGGACTCCTGTGATGATTGA 2993.

QY 3045 ATCTGTTCAACTGAAAAACCTC 3065

Db 2994 CTCAGTCCGCTGAAGACATC 3014

RESULT 5

AX127726

LOCUS AX127726 3331 bp DNA linear PAT 15-MAY-2001

DEFINITION Sequence 1 from Patent WO0131003.

ACCESSION AX127726

VERSION AX127726.1 GI:14134373

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 3331)

AUTHORS Delneste,Y., Magistrelli,G., Jeannin,P. and Bonnefoy,J.Y.

TITLE Cloning, expression and characterisation of a gene expressed in tumour cells and involved in the regulation of the immune response

JOURNAL Patent: WO 0131003-A 1 03-MAY-2001;

PIERRE FABRE MEDICAMENT (FR)

FEATURES

source

1. .3331

/organism="Homo sapiens"

/db\_xref="taxon:9606"

45. .3086

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAC38907.1"

/db\_xref="GI:14134374"

/translation="MAEPGSHSHLSARVRGRTERRRIPRLWRLLWAGTAFQVTQGTGP

ELHACKESEHYHEYTACDSTGSRWRVAVPHTPGLCTSLDPVKGTESFSCNAGEFLD

MKDQCKPCAEGRYSIGTGIRFDEWDELPHGFASLSANMELDDSAESTGCTSSKVV

PRGDIASNTDECTATLMYAVNLKQSGTVNFEYYPDSSIIIEFFVQNDQCPNADDS

RWMKTTKEGWEFHSVELNRGNVLYWRRTAFSVWTKVPKPLVRNIAITGVAYTSECF

PKCPGTYADKQGSFCKLCPANSYSNKGETSCHQCDPDKYSEKSSSCNVRPACTDKD

YFYHTACDANGETQLMYKWAKPKICSEDLGAVKLPASGVKTHCPNCPGFFKTNNS

TCQPCYPYNSGSDCTRCPAGTEPAVGFEYKWNLTLPNTMETTVLSGINFEYKGMTG

WEVAGDHIYTAGASDNDFMILTLPVPGFRPQSVMDTENKEVARITFVFETLCSVN

CELYFMVGNRSRNTPVETWKGSKGQSYTIIIEENTTSFTWAFORTTFHEASRKYT

NDVAKIYSINVTNVMNGVASYCRPCALEASDVSSCTSCPAGYYIDRDSGTCHSCPNN

TILKAHQPYGVQACVPCGPGTKNNKIHSLCYNDCTFSRNTPTRTFNYNFSALANTVTL

AGPSFTSKGLKYFHHFTLSLCGNQGRKMSVCTDNVTDLRIPEGESGFSKSITAYVCQ

AVIIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSPAEFLHLESIGIPDVIFF

YRSNDVTQSCSSGRSTTIRVRCSPQKTPVGSLLLPCTSDGTCDCGNFHLWESAAC

PLCSVADYRAIVSSCVAGIQKTTYVWREPKLCSGGISLPEQRTICKTIDFWLKVGIS

AGCTAILLTVLTCYFWKKNQKLEYKYSKLVMNATLKDCLDPAADSCAIMEGEDVEDD

LIFTSKSLFLGKIKSFTSKRTPDGFDSVPLKTSSGGPDMDL"

BASE COUNT 822 a 945 c 813 g 751 t

ORIGIN

Query Match 27.8%; Score 989.8; DB 6; Length 3331;

Best Local Similarity 60.6%; Pred. No. 1.8e-221;

Matches 1757; Conservative 0; Mismatches 1117; Indels 27; Gaps 7;

QY 171 GCTTCCTCCTTGCCAGGAGAGAAAGATTATCAGTTTGAATATACGGAATGTGATAGCAGTGG 230

Db 179 GCTTCATGCCTGCAAAAGAGTCTGAGTACCACCTATGAGTACACGGCGTGTGACAGCACGGG 238

QY 231 CTCCAGGTGGAGAGTTGCCATTCCAAATTCCTGCAGTGGACTGCTCTGGCCTGCCTGACCC 290

Db 239 TTCCAGGTGGAGGTGCGCGTGCCGCATACCCCGGCCCTGTGCACACGCGCTGCCTGACCC 298

QY 291 AGTGAGAGGCAAGAATGCACCTTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAA 350

Db 299 CGTCAAGGGCACCGAGTGTCTCTCTGCAAGCCGGGGAGTTTCTGGATATGAAGGA 358

QY 351 CCAGGTATGCAGTAAAGTGTGGTGAAGGCACCTATTCTTGGCAGTGGCATCAAATTTGA 410

Db 359 CCAGTCATGTAGCCATCGCTGAGGGCGGCTACTCCTCGGCACAGGCATTCGGTTTGA 418

QY 411 TGAATGGGATGAATGCCGGCAGGATTTTCTAAACATCGCAACATTTCATGGACACTGTGGT 470

Db 419 TGAGTGGGATGAGTGCCTCATGGCTTTGCCAGCCTCTCAGCCAAACATGGAGCTGGATGA 478

QY 471 GGGCCCTTCTGACAGCAGGCGGCTGTAACAACTCTTCTTGGATCCCTCGTGGAAA 530

Db 479 CAGTGTCTGCTGACTCCA--CCGGGAACCTGTACTTCGTCCAAGTGGGTCCCCGGGCGGA 535

QY 531 CTACATAGAATCTAATCGTGACTGCACCGGTGTCTTTGATCTATGCTGTGCACCTTAA 590

Db 536 CTACATCGCCTCCAACACGGACGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 595

QY 591 GAAGTCAGGCTATGTCTTCTTTGAGTACCAGTATGTCGACAACAACATCTCTCTTTGAGTT 650

Db 596 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 655

QY 651 CTTTATTCAAATGATCAGTGCCAGGAGATGGACACCACCTGACAAGTGGGTAAAACT 710

Db 656 TTTCTGTTCAGAAATGACCAGTGCCAGCCCAATGCAGATGACTC--CAGTGGATGAAGAC 712

QY 711 TACAGACAATGGAGAATGGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAAACATACT 770

Db 713 CACAGAGAA--AGGATGGGAATTCCACACTGTGGAGCTAAATCGAGGCAATAATGTCCT 769

QY 771 CTACTGGAGAACTACAGGCATCCTTATGGTTCTAAGGGGTCAAGCCTGTGCTGGTAAA 830

Db 770 CTATTGGAGAACCAACAGCCTTCTCAGTATGGACCAAAAGTACCCCAAGCCTGTGCTGGTGAG 829

QY 831 AAATATCACAATTGAAGGGGTGGCGGTACACATCAGAAATGTTTTCTTTCGCAAGCCAGGCAC 890

Db 830 AAACATTGCCATAACAGGGGTGGCCTACACTTCAGAAATGCTTCCCTTGCAAAACCTGGCAC 889

QY 891 ATTCAGCAACAACACAGGTTTCATTCAACTGCCAGGTGTGTCACAGAAACACCTATTCTGA 950

Db 890 GTATGCAGACAAGCAGGGCTCCTCTTTCTGCAAACTTTGCCACGCAACTCTTATTCAAA 949

QY 951 GAAAGGAGCAAAAGAAATGTATAAGGTGTAA--AGACGACTCTCAATTTTCAGGATCCAG 1007

Db 950 TAAAGGAGAAACTTCTTGCCACCAGTGTGACCCCTGACAAATACTCAGAGAAAAGGATCTTC 1009

QY 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCCACAAAAGACTATTTCCAGATCCATCTCCATG 1067

Db 1010 TTCTCTGTAACTGCGGCCAGCTTGCACAGACAAGAGATTATTTCTACACACACACGCGCCTG 1069

QY	1068	1068	TGATGAAGAAGGAAAGACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGGA	1127
DB	1070	CGATGCCAACGGAGAGACACAACTCATGTACAATGGGCCAAGCCGAAAATCTGTAGCGA	1129	1127
QY	1128	GGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAAGAAGGATTGTCGCCCTTG	1187	1129
DB	1130	GGACCTTGAGGGGCAGTGAAGCTGCCTGCCTCTGGTGTGAAGACCCACTGCCCAACCTTG	1189	1187
QY	1188	CAACCCCTGGATTTTATAACAATGGATCATCTTGTGCCATCCCTGTCTCTGGAACATTT	1247	1189
DB	1190	CAACCCAGGCTTCTTCAAAACCAACAACAGACACCTGCCAGCCCTGCCCATATGGTCCCTA	1249	1247
QY	1248	TTCAGATGGAACCAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTT	1307	1249
DB	1250	CTCCAATGG--CTCAGACTGTACCCGCTGCCCTGCAGGACTGAACCTGCTGTGGGATT	1306	1307
QY	1308	TGAATATAAATGGTGGAAATGTCTTCTTGGCAACATGAAAACTTCCCTGCTTCAATGTTGG	1367	1306
DB	1307	TGAATACAAATGGTGGAAACACGCTGCCACAAACATGGAACGACCGTTCTCAGTGGGAT	1366	1367
QY	1368	GAATTCAAAGTGCAGTGAATGAATGGTTGGAGGTGGCTGGAGATCATATCCAGAGTGG	1427	1366
DB	1367	CAACTTCGAGTACAAGGCATGACAGGCTGGAGGTGGCTGGTGATCACATTTACACAGC	1426	1427
QY	1428	GGCTGGAGTCTTGACAAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACC	1487	1426
DB	1427	TGCTGGAGCCTCAGACAAATGACTTCATGATTCTCACTCTGGTTGTGCCAGGATTTAGACC	1486	1487
QY	1488	ACCAACATCTATGACTGGAGCCAC--GGGTTCTGAACTAGGAAGAATAACATTTGTCTTT	1544	1486
DB	1487	TCCGCAGTCGGTGATGGCAGACACAGAGAATAAAGAGGTGGCCAGAATCACATTTGTCTT	1546	1544
QY	1545	TGAGACCCCTGTGTGACTGACTGTGTTTGTACTTCATGCTGGTGGATATTAATAGAAAAAG	1604	1546
DB	1547	TGAGACCCCTGTGTGAACTGTGAGCTCTACTTCTCATGTTGGTGGTGTAATTTCTAGGAC	1606	1604
QY	1605	TACAAATGTGTAGAATCGTGGGTGGAAACCAAGAAAAACAAGCTTACACCCCATATCAT	1664	1606
DB	1607	CAACACTCCTGTGGAGACGTGGAAGGTTCCAAAGGCCAAACAGTCTTATACCTACATCAT	1666	1664
QY	1665	CTTCAAGAATGCAACTTTTACATTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGA	1724	1666
DB	1667	TGAGGAGAACACTACCACGAGCTTCACCTGGGCCTTCCAGAGGACCACCTTTTCATGAGGC	1726	1724
QY	1725	TAATAGACGGTTCTATCAATGACATGGTGAAGATTTTATCTATCACAGCCCACTAATGCAGT	1784	1726
DB	1727	AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCACCAATGTTAT	1786	1784
QY	1785	TGATGGGGTGGCGTCTCATGCCGTGCCCTGTGCCCTCGGTTCTGAACAGTTCGGTTTCATC	1844	1786
DB	1787	GAATGGCGTGGCCTCCTACTCCTGCTGCCCTGTGCCCTAGAACGCTCTGATGTGGCTCCTC	1846	1844
QY	1845	GTGTGTCCCCCTGCCCTCCAGGCCACTACATTTGAGAAAGAAACCAACCAAGTCAAGGAATG	1904	1846
DB	1847	CTGCACCTCTTGTCTGCTGCTTACTATATTGACCCGAGATTCAGSAACCTGCCACTCCTG	1906	1904
QY	1905	TCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTCCCATG	1964	1906
DB	1907	CCCCCTTAACACAAATCTGAAAGCCCAACGACCTTATGGTGTCCAGGCTGTGTGCCCTG	1966	1964
QY	1965	CGGGCCTGGGAGTAAAAACAATCAGGACCATTCCGTTTGGTATAGTGACTGCTTTTCTA	2024	1966
DB	1967	TGGTCCAGGGACCAAGAACAACAGATCCACTCTCTGTGCTACAAATGATTGCACCTTCTC	2026	2024
QY	2025	CCATGAAAAAGAAAATCAGATTTTGCACATATGACTTTTAGCAACCTCAGCAGTGTGGGCTC	2084	2026
DB	2027	ACGCAACACTCCAACCAAGGACTTTCAACTACAACCTTCTCCGCTTTGGCAACACACCGTCAC	2086	2084
QY	2085	ATTAATGAATGGCCCCCAGCTTCACCTCCAAGGAACAAAATACCTTCCATTTCTTCAATAT	2144	2086
DB	2087	TCCTTGCTGGAGGCCCAAGCTTCACCTTCCAAGGGTTGAAATACTTCCCATCACTTACCCT	2146	2144
QY	2145	CASCTTATGTGGGCATGAGGGGAAGATGGCTCTCTGTACCAACAATATACAGACTT	2204	2146

[illegible]

## RESULT 6

AX463904

LOCUS

LOCUS	AX463904	3501 bp
DEFINITION	Sequence 37 from patent WO0140466.	

DEFINITION  
ACCESSION  
AX463904

ACCESSION	AX403504	GI:21898969
VERSION	AX463904.1	

VERSION  
KEYWORDS

KEYWORDS . human.  
SOURCE .  
ORGANISM Homo sapiens

linear PAT 16-JUL-2002

3507 bp DNA

3507 bp DNA

PAT 16-JUL-2002





Qy	2025	CCATGAAAAAGAAATCAGATTTTGCACATATGACTTTTAGCAACCTTCAGCAGTGTGGGCTC	2084
Db	2115	ACGCAACACTCCAACACGAGACTTTCAACTACAACTTCTCCGCTTTGGCAACACACCGTCAC	2174
Qy	2085	ATTAATGAATGGCCCCAGCTTCACCTCCAAAGGAACAATACTATTCCATTTCTTCAATAT	2144
Db	2175	TCATTGCTGGAGGCCCAAGCTTCACCTCCAAAGGGTTGAAATACTTCCATCACTTTACCCCT	2234
Qy	2145	CAGTTTATGTGGGCATGAGGGGAAGAGATGGCTCTCTGTACCAACAATATAACAGACTT	2204
Db	2235	CAGTCTCTGTGGAACACGAGGTAGGAAAATGTCTGTGTGACCCGACAATGTCACTGACCT	2294
Qy	2205	TACAGTAAAGAAATAGTGGCAGGGTCAGATGATTACACAAATTTGGTAGGGGCATTTGT	2264
Db	2295	CCGGATTCCCTGAGGGTGAGTCAGG-----TTCTCCAAATCTATCACAGCCTACCT	2345
Qy	2265	ATGCCAGTCAACAATATTCCCTTCTGAAAGTAAGGGTTTCCGAGCAGCCTTATCATCACA	2324
Db	2346	CTGCCAGGCAGTCATCATCCCCCAGAGGTGACAGGCTACAAGGCCGGGTTTCCTCACA	2405
Qy	2325	ATCCATCATTTCTGGCAGATACATTTCATAGGAGTCACAGTTGAAACCACATTTGAAAATAT	2384
Db	2406	GCCTGTAGCCTTGTCTGATCGACTTATTGGGGTGACAACAGATATGACTCTGGATGGAAT	2465
Qy	2385	TAATATAAAGAAAGATATGTTCCCGAGTTCCAACAAGCCCAATACCAGATGTGCATTTCTT	2444
Db	2466	CACCTCCCCAGCTGAACCTTTTCCACCTGGAGTCTCTTGGGAATACCCGACGTGATCTTCTT	2525
Qy	2445	TTATAAGTCTTCTACAGCAACAACATCTTGTATTAAATGGCCGATCAACTGCTGTGAAT	2504
Db	2526	TTATAGTCCAATGATGTACCCAGTCTCTGCAGTTCTGGGAGATCAACCACCATCCGCGT	2585
Qy	2505	GAGTGTAAATCTACTAAATCTGGAGCAGGAGTGATTTTCACTCCCCCAGCAAGTGCCCCAGC	2564
Db	2586	CAGGTGCAGTCCACAGAAAACGTGCCCTGGAAAGTTTGTGCTGCTGCCAGGAACGTGCTCAG	2645
Qy	2565	AGGTACCTGTGATGGGTGTACGTTCTATTTCCTGTGGGAGAGTGTGAAGCTTGCCCTCT	2624
Db	2646	TGGGACCTGTGATGGCTGCAACTTCCACTTCTGTGGGAGAGCGCGGTGCTTGCCCCGT	2705
Qy	2625	GTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCTGCAAGAGAGAGGATTTCAGGAAC	2684
Db	2706	CTGCTCAGTGGCTGACTACCATGCTATCGTCAGCAGCTGTGTGGCTGGGATCCAGANGAC	2765
Qy	2685	CTTGATGTGTGGAATGAACCTAAATGGTGCATTAAAGGAATTTCTTTGCCCTGAGAAAAA	2744
Db	2766	TACTTACGTGTGNCGAGAACCCCAAGCTATGCTCTGGTGGCATTTCTCTGCCTGAGCAGAG	2825
Qy	2745	GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGGTGGGAGCCGGTGTGGGAGCTTT	2804
Db	2826	AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG	2885
Qy	2805	TACTGCCGTTTGTGTTGGCTCTGACCTGCTACTTCTGGAAAAAGAATCAAAAACTGGA	2864
Db	2886	TACTGCCATCCTGCTCACCCGTCCTTGACCTGCTACTTTTGGAAAAAGAATCAAAAACTAGA	2945
Qy	2865	ATACAAATATTCCAAGTTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCCGGCTGC	2924
Db	2946	GTACAAGTACTCCAAGCTGGTGATGAATGCTACTCTCAAGGACTGTGACCTGCCAGCAGC	3005
Qy	2925	AGACAGTTGTGCTATCATGGAAGGAGAAGATAATGAAGAGGAAGTTGTATATTCCAATAA	2984
Db	3006	TGACAGCTGCGCCCATCATGGAAGCGGAGGATGTAGAGGACGACCTCATCTTTACCAGCAA	3065
Qy	2985	ACAGTCTACTAGGAAAACCTCAATCTTTGGCAACCAAGGAAAAAGAAGACCATTTTGA	3044
Db	3066	GAAGTCACTTTTGGGAAGATCAAAATCATTTACCTCCAAGAGGACTCCTGTATGGATTGA	3125
Qy	3045	ATCTGTCAACTGAAAACCTC	3065
Db	3126	CTCAGTGGCGCTGAAGACATC	3146

RESULT 7	AX127742	AX127742	2610 bp	DNA	linear	PAT 15-MAY-2001
LOCUS	Sequence 17 from Patent WO0131003.					
DEFINITION	AX127742					
ACCESSION	AX127742					
VERSION	AX127742.1	GI:14134389				
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 2610)					
AUTHORS	Delneste,Y., Magistrelli,G., Jeannin,P. and Bonnefoy,J.Y.					
TITLE	Cloning, expression and characterisation of a gene expressed in					
	tumour cells and involved in the regulation of the immune response					
JOURNAL	Patent: WO 0131003-A 17 03-MAY-2001;					
	PIERRE FABRE MEDICAMENT (FR)					
FEATURES	Location/Qualifiers					
source	1..2610					
	/organism="Homo saplens"					
	/db_xref="taxon:9606"					
CDS	<1..>2610					
	/note="unnamed protein product"					
	/codon_start=1					
	/protein_id="CAC38913.1"					
	/db_xref="GI:14134390"					
	/translation="TGPELHACKESEYHYEYACDSTGSRWRVAVPHTPGLCTSLPDP					
	VKTECSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMEL					
	DDSAAESTGNCTSSKWWPRGDYIASNTDECTATLMYAVNLKQSGTVNEYYPDSSII					
	FEFFVONDQCOPNADDSRWMTTEKGWEFHSVELNRGNLVYRWTAFSVMTKVPKPV					
	LVRNIAITGVAYTSECPCPKPGTYADKQGSFCKLCPANSYSNKGFTSCHQCDPKYS					
	EKSSSCNVRPACTDKDYFYTHTACDANGETQLMYKWAKPKICSEDLEGAVKLPASGV					
	KTHCPPCNPGFKTNNSTCQPCPYGPYSNGSDCTRCPAGTEPAVGFYKWNWTLPTNM					
	ETTVLSGINFEYKGMTGWEVAGDHIYTAGASDNDFMILTLLVVPGRPPQSVMDTEN					
	KEVARITFEFETLCSVNCELYFMVGVNSRNTPVETWKGSKGKQSYTIIIEENTTSF					
	TWAFQRTTFHEASRKYTNDVAKIYSINVTNVMNGVASYCRPCALEASDVSSCTSCPA					
	GYVIDRDSGTCHSCPNTILKAHQPYGVQACVPCGPGTKNNKIHSCLVNDCTFSRNTP					
	TRTFYNFSALANTVTLAGGPSFTSKGLYFHHFTLSLCGNQGRKMSVCTIDNIDLRI					
	PEGESGFSKSITAYVCOAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSP					
	AELFHLESIGIPDIVIFFYRSNDVTQSCSRSTTIRVRCSPQKTVPGSLLPLPGTCSDG					
	TCDCGNFHFLEWESAAACPLCSVADYRAIVSSCVAGIQKTTYVWREPKLCSGGISLPEQ					
	RVTICKTIDFWLKV"					
BASE COUNT	634 a	744 c	644 g	588 t		
ORIGIN						
Query Match	24.2%;	Score	860.8;	DB 6;	Length	2610;
Best Local Similarity	60.0%;	Pred.	No. 3.5e-191;			
Matches	1571;	Conservative	0;	Mismatches	1022;	Indels 27; Gaps 7;
Qy	171	GC	TTCCCTCCTTGCCAGGAGAGAAAGATTTATCAC	TTTGAATATACGGAATGTGATAGCAGTGG	230	
Db	12	GC	TTTCATGCCCTGCAAAAGAGTCTGAGTACCAC	TATGAGTACACGCGGTGTGACAGCACGGG	71	
Qy	231	CT	CCAGGTGGAGAGTTGCCATTTCCAAATTTCTGC	CAGTGGACTGCTCTGGCCCTGCCCTGACCC	290	
Db	72	TT	CCAGGTGGAGGGTCCGCGTCCCGCATACCCCG	GGCCTGTGCACACGCTGCCCTGACCC	131	
Qy	291	AG	TGAGAGGCAAAAGATGCAC	TTTCTCCTGTGCTTCTTGGAGAGTATCTAGA	AAATGAAGAA	350
Db	132	CG	TCAAGGACCGAGTGTCTCTCTCTGCAACG	CCCGGGAGTTCCTGGATATGAAGGA	191	
Qy	351	CC	AGGTATGCAGTAAGTGTGGTGAAGGCACCT	ATTCCCTGGGCAGTGGCATCAAAATTTGA	410	
Db	192	CC	AGTCATGTAAGCCATCGCTGAGGCGCGCT	ACTCCCTCGGCACAGGCATTCGGTTTGA	251	
Qy	411	TG	AATGGGATGAATTGCCGGCAGGATTTTCT	TAACATCGCAACATTCATGGACACTGTGGT	470	
Db	252	TG	AGTGGGATGAGCTGCCCCATGGCTTTGCC	AGCCTCTCAGCCCAACATGGAGCTGGATGA	311	
Qy	471	GG	CCCTTCTTGACAGCAGGCCAGCGGCTGT	AACAAC	TCTTCTTGGATCCCTCGTGGAAA	530
Db	312	CAG	TGCTGAGTCCA--	CCGGGA	ACTGTACTTCTGTC	CAAGTGGGTTC
						CCCCGGGGCGGA



QY 531 CTACATAGAAATCTAAATCGTGATGACTGCACGGGTGCTTTTGATCTATGCTGTGCACCTTAA 590  
Db 369 CTACATCGCCTCCAACACGCGAGAAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 428  
QY 591 GAAGTCAGGCTATGTCTTCTTTGAGTACCAGTATGTCGACAACAACATCTTCTTTGAGTT 650  
Db 429 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 488  
QY 651 CTTTATTCAAAATGATCAGTCCAGGAGATGGACACCCACTGCACAAGTGGGTAAACT 710  
Db 489 TTTTCGTTCAGAATGACCAGTCCAGGCCAATGCAGATGACTC---CAGGTGGATGAAC 545  
QY 711 TACAGACAATGGAGAAATGGGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATACT 770  
Db 546 CACAGAGAA---AGGATGGGAATCCACAGTGTGGAGCTAAATCGAGGCAATAATGTCT 602  
QY 771 CTACTGGAGAACTACAGGCATCCTTATGGGTTCTAAGCGGTCAAGCCTGTGCTGTTAAA 830  
Db 603 CTATTGGGAACCCACAGCCCTTCTCAGTATGGACCAAGTACCCAAAGCTACCCAAAGCCTGTGCTGGTGAG 662  
QY 831 AAATATCACAATTAAGGGGTGGCTGACATCAGAAATGTTTTCCTTGTCAAGCCAGGCAC 890  
Db 663 AAACATTGCCATAAACAGGGGTGGCTTACACTTCAGAATGCTTCCCTTGCAACCTTGGCAC 722  
QY 891 ATTACGCAACAAACCAGGTTCAITCAACTGCCAGGTGTGTCCAGAAACACCTATTCTGA 950  
Db 723 GTATGCAGACAAGCAGGGCTCCTCTTTCTGCAAACTTTGCCACAGCCAACTCTTATTCAA 782  
QY 951 GAAAGGAGCCAAAGAATGTATAAGGTGTAA---AGACGACTCTCAATTTTCAGGATCCAG 1007  
Db 783 TAAAGGAGAAACTTCTTGCCACCAGTGTGACCCTTGACAAATACTCAGAGAAAGGATCTTC 842  
QY 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATACTCCATG 1067  
Db 843 TTCTGTGTACGTGCGCCAGCTTGCACAGACAAAAGATTATTTCTACACACACACGGCCTG 902  
QY 1068 TGATGAAGAGAAAGACACAGATAATGTACAAAGTGGATAGAGCCCAAAATCTGCCGGGA 1127  
Db 903 CGATGCCAACGGAGAGACACAACATCATGTACAAATGGGCCAAGCCGCAAAATCTGTAGCGA 962  
QY 1128 GGATCTCACAGATGCTATTAGATTGCCCCCTTCTTGCCATCCCTGFCCTCCTGGAACATT 1247  
Db 1023 CAACCCAGGCTTCTTCAAAACCAACAACAGCACCTGCCAGCCCTGCCATATGTTCCCTA 1082  
QY 1248 TTCAGATGGAACCAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTT 1307  
Db 1083 CTCCAATGG---CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGATT 1139  
QY 1308 TGAATATAAATGGTGGAAATGTCCTTTCGCAACATGAAAACTTCTCTGCTTCAATGTTGG 1367  
Db 1140 TGAATACAAATGGTGGAAACACGCTGCCCAACAACATGGAAACGACCGTCTCAGTGGGAT 1199  
QY 1368 GAATTCAAAGTCGATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427  
Db 1200 CAACTTCGAGTACAAGGGCATGACAGGCTGGGAGGTGGCTGGTGATCACATTTACACAGC 1259  
QY 1428 GGCTGGAGGTTCTGACAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAACC 1487  
Db 1260 TGCTGGAGCCTCAGACAATGACTTCATGATTTCTCACTCTGTTGTGCGCAGGATTTAGACC 1319  
QY 1488 ACCAACATCTATGACTGGAGCCAC---GGGTTCTGAAC TAGGAAGAAATAACATTTGTCTT 1544  
Db 1320 TCCGCAGTCGGTGGTGGACACACAGAGAATAAAGAGGTGGCCAGAAATCACATTTGTCTT 1379  
QY 1545 TGAGACCCCTCTGTTACAGCTGACTGTGTTTTGTACTTCATGGTGGATATTAATAGAAAAAG 1604  
Db 1380 TGAGACCCCTCTGTTCTGTGAACCTGTGAGCTCTACTTCAATGGTGGGTGTTGAATTTAGGAC 1439  
QY 1605 TACAAATGTGGTAGAATCGTGGGTGGAAACCAAGAAAAACAAGCTTACACCCCATATCAT 1664

Db 1440 CAACACTCCTGTGGAGACCTGGAAAGGTTCCAAAGCGCAACACAGTCCCTATACCTACATCAT 1499  
QY 1665 CTTCAAGAAATGCAACTTTTACATTTACATGGGCATTCACAGAGAACTAATCAGGGTCAAGA 1724  
Db 1500 TGAGGAGAACACTACCAGAGCTTCACCTGGGCCCTCCAGAGGACCACCTTTTCATGAGGC 1559  
QY 1725 TAATAGACGGTTCAATCAATGACATGGTGAAGATTTATCTATCAGAGCCACTAATGTCAGT 1784  
Db 1560 AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCAACCAATGTTAT 1619  
QY 1785 TGATGGGTGGCGTCCCTCATGCGCTGCCCTGTGCCCCCTCGGTTCTGAAACAGTCGGGTTTCATC 1844  
Db 1620 GAATGGCGTGGCCTCCTACTGCTGCCCTGCCCTAGAACCTCTGTATGTGGGCTCCTC 1679  
QY 1845 GTGTGTCCTCCCTGCCCTCCAGGCCACTACATTTGAGAAAAAGAAACCAACAGTGCAGGAATG 1904  
Db 1680 CTGCACCTCTTGTCTGCTGCTGTACTATATTGACCGAGATTCAGGAACCTGCCACTCCTG 1739  
QY 1905 TCCACCTGACACCTTACCCTGCCATACATCAGGTCTATGSCAAAGAGGCTTGATTCCCATG 1964  
Db 1740 CCCCCCTAACACAATTCTGAAAGCCCCACCAGCCTTATGGTGTCCAGGCCCTGTGTGCCCTG 1799  
QY 1965 CGGGCCTGGGAGTAAAAACAATCAGGACCATTCCGGTTTTCATATAGTGAAGTCTTTTCTA 2024  
Db 1800 TGGTCCAGGGACCAAGAACACAAGATCCACTCTCTGTGCTACAAATGATTGCACCTTCTC 1859  
QY 2025 CCATGAAAAAGAAATCAGATTTTGCACTATGACTTTAGCAACCTCAGCAGTGTGGGCTC 2084  
Db 1860 ACGCAACACTCCAACCCAGGACTTCAACTACAACCTTCTCGGCTTGGCAACACCCGTCAC 1919  
QY 2085 ATTAATGAATGGCCCCAGCTTCACCTCCAAGGAACAATAACTTCCATTCTTCAATAT 2144  
Db 1920 TCTTGTCTGGAGGGCCAAAGCTTCACTTCCAAGGGTTGAAATACTTCCATCACTTTACCCT 1979  
QY 2145 CAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATAAACAGACTT 2204  
Db 1980 CAGTCTCTGTGGAAACCAGGGTAGGAAAAATGTCTGTGTGCACCGACAATGTCACTGACCT 2039  
QY 2205 TACAGTAAAAAGAAATAGTGGCAGGTCAGATGATTACACAAATTTGGTAGGGCATTTTGT 2264  
Db 2040 CCGGATTCCTGAGGTGAGTCAGGGTT-----CTCCAATCTATCACAGCCTACGT 2090  
QY 2265 ATGCCAGTCAACAATATTTCCTTCTGAAAGTAAGGGTTTCCGAGCAGCCTTATCATCACA 2324  
Db 2091 CTGCCAGCAGTCAATATCCCCCAGAGGTCACAGGCTACAGGCCGGGTTTCTCTCACA 2150  
QY 2325 ATCCATCATCTTGGCAGATACATTCATAGGAGTCACAGTTGAAACCAACATTTGAAAAAT 2384  
Db 2151 GCCTGTACGCTTGTGTATCGACTTATTTGGGTTGACAACAGATGACTCTGGATGGAAT 2210  
QY 2385 TAATAAAGAAAGATATGTTCCCACTTCCAAGTTCCAACAAGCCAAATACCAGATGTGCATTTCTT 2444  
Db 2211 CACCTCCCAGCTGAACCTTTTCCACCTGGAGTCTTTGGGAATACCGGACGTGATCTTCTT 2270  
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAAATGGCCGATCAAACTGCTGTGAAAAAT 2504  
Db 2271 TTATAGGTCCAATGATGTGACCCAGTCTTGCAGTTCTGGGAGATCAACCAACCATCCGCGT 2330  
QY 2505 GAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCCAGCAAGTGCCCCAGC 2564  
Db 2331 CAGGTGCAGTCCACAGAAAACTGTCCCTGGAAAGTTTGTCTGCCAGGAACGTGCTCAGA 2390  
QY 2565 AGTACCTGTGATGGGTGTACGTTCTATTTCCTGTGGGAGAGTGCTGAAGCTTGCCCTCT 2624  
Db 2391 TGGGACCTGTGATGGCTGCAACTTCCACTTCTGTGGGAGAGCGCGGCTGCTTGCCCGCT 2450  
QY 2625 GTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCTGCAAGAGAGGATTTTCAGGAAC 2684  
Db 2451 CTGCTCAGTGGCTGACTACCGTGTATCGTCAGCAGCTGTGTGGCTGGGATCCAGAAGAC 2510  
QY 2685 CTTGTATGTGGGAATGAACCTTAAATGCTGCATTAAGGAATTTCTTTGCTGTGAAAAA 2744

Db 2511 TACTTACGTGGCGAGAACCCAAAGCTATGCTCTGGTGGCATTTCTCTGCCTGAGCAGAG 2570

QY 2745 GTTGGCAACTGTGAACACGGTTGACTTTTGGCTGAAGGTG 2784

Db 2571 AGTCACCATCTGCAAAACCATAGATTCTCGCTGAAAGTG 2610

RESULT 8

AX127740

LOCUS AX127740 2733 bp DNA linear PAT 15-MAY-2001

DEFINITION Sequence 15 from Patent WO0131003.

ACCESSION AX127740

VERSION AX127740.1 GI:14134387

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2733)

AUTHORS Delneste,Y., Magistrelli,G., Jeannin,P. and Bonnefoy,J.Y.

TITLE Cloning, expression and characterisation of a gene expressed in tumour cells and involved in the regulation of the immune response

JOURNAL Patent: WO 0131003-A 15 03-MAY-2001;

PIERRE FABRE MEDICAMENT (FR)

FEATURES

source Location/Qualifiers

1..2733

/organism="Homo sapiens"

/db\_xref="taxon:9606"

1..>2733

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAC38912.1"

/db\_xref="GI:14134388"

/translation="MAEPGSHSHLSARVRGRTERRIPRLWRLLLWAGTAFQVQTGTP ELHACKSEYHYEYACDSTGSRWRVAVPHTPLCLTSLDPVKGTCEFSFNAGEFLD MKDQSCKPCAEGRYSLGTGIRFEWDELPHGFASLSANMELDDSAESTGNTSSKWV PRGDYIASNTDECTALMYAVNLKQSGTVNFYYPDSSIIFFEYVQNDQCPNADDS RWMKTEKGWEFHSVELNRGNVLYWRTTAFSVWTKVPKPLVLRNIAITGVAYTSECF PCKPGTYADKQGSFCKLCPANSYSNKGTSCHODPDPKYSEKGSSSCNVRPACTDKD YFYTHTACDANGETQLMYKWKAPKICSEDLLEGAVKLPAAGVKTHCPCPCNPFKTNNS TCQCPYPGYSNGSDCTRCPCAGTEPAVGFEYKWNLTPTNMTTVLSGINFEYKGMTG WEVAGDHIIYTAAGASDNDFMILTLLVVPGRFPQSQVMDTENKEVARITTFETLCSVN CELYFMVGVNSRTNTPVETWKGSKGQSYTIIEENTTTSTWAFORTTTFHEASRKYT NDVAKIYSINVTNVMNGVASYCRPCALEASDVSSCTSCPAGYYIDRSGTCHSCPPN TILKAHPYGVQACVPCPGTKNNKIHSCLYNDCTFSRNTPTRTFNYNFESALANTVL AGPSFTSKGLKYFHHTLSLSCGNQGRKMSVCTDNVTLRIPEGESGFSKSITAYVCQ AVIIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSAPAEFLHLESIGIPDVIFE YRSNDVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCSDGTCDCGNFHFIMESAAAC PLCSVADYRAIVSSCVAGIQKTTYVWREPKLCSGGISLPEQRVTICKTIDFWLKV"

BASE COUNT 656 a 783 c 687 g 607 t

ORIGIN

Query Match 24.2%; Score 860.8; DB 6; Length 2733;

Best Local Similarity 60.0%; Pred. No. 3.5e-191;

Matches 1571; Conservative 0; Mismatches 1022; Indels 27; Gaps 7;

QY 171 GCTTCCTCCTGCCAGGAGAAAGATTATCACTTTTGAATATACGGAATGTGATAGCAGTGG 230

Db 135 GCTTCATGCGCTGCAAAAGAGTCTGAGTACCACCTATGAGTACACGGCGTGTGACAGCACGGG 194

QY 231 CTCACAGTGGAGAGTTGCCATTCCTGCTGCTGCTGAGTGGAGTGTGATGATATAGAAATGAAGAA 350

Db 255 CGTCAAGGSCACCGAGTGCTCCTCTCTGCAACGCCGGGGAGTTTCTGGATATGAAGGA 314

QY 351 CCAGGTATGCAGTAAGTGTGGTGAAGGCACCTATTCTCTTGGGCAGTGGGCATCAAAATTTGA 410

Db 315 CCAGTCATGTAAGCCATGCGCTGAGGGCGGCTACTCCCTCGGCACAGGCATTCGGTTTGA 374

QY 411 TGAATGGGATGAATTGCCGGCAGGATTCTTAACATCGCAACATTCATGGACACTGTGGT 470

Db 375 TGAGTGGGATGAGCTGCCCATGGCTTTGCCAGCCTCTCAGCCAACATGGAGCTGGATGA 434

QY 471 GGGCCTTCTGACAGCAGGCCAGACGGCTGTAAACAACTCCTTCTGGATCCCTCGTGGAAA 530

Db 435 CAGTGTCTGCTGAGTCCA--CCGGGAACGTGTACTTCGTCAAGTGGGTTCCTCCGGGGCGA 491

QY 531 CTACATAGAATCTAATCGTGTGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAA 590

Db 492 CTACATCGCCTCCAACACACGGACGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 551

QY 591 GAAGTCAGGCTATGTCTTTTGTAGTACCAGTATGTGACAAACAACATCTTCTTTGAGTT 650

Db 552 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 611

QY 651 CTTTATTCAAAATGATCAGTGGCAGGAGATGGACACCCACACTGACAAGTGGGTAAAACT 710

Db 612 TTTCTGTTCAGAAATGACCAGTGGCAGCCCAATGCAGATGACTC--CAGGTGGATGAAGAC 668

QY 711 TACAGACAATGGAGAATGGGCTCTCTATTCTGTAAATGCTGAATCAGGCACAAACATPACT 770

Db 669 CACAGAGAA--AGGATGGGAATTCACACAGTGTGAGCTGTAATCGAGGCAATAATGTCTT 725

QY 771 CTACTGGAGAACTACAGGCACTCTTATGGTTCCTAAGGGGTCAAGCCCTGTGCTGGTAAA 830

Db 726 CTATTGGAGAACACACAGCCTTCTCAGTATGGACCAAAAGTACCCTGCTGCTGGTGAG 785

QY 831 AAATATCACAAATTTGAAGGGTGGCGGTACACATCAGAAATCTTTCTTGAAGCCAGGCAC 890

Db 786 AAACATTTGCCATAAACAGGGGTGGCTTACACTTCAGAAATGCTTCCCCTGCMAAACCTGGCAC 845

QY 891 ATTACAGCAACAAACCAGGTTTCATTCAACTGCCAGGTGTGCCAGAAACACACCTATTCTGA 950

Db 846 GTATGCAGACAAGCAGGGCTCCTCTTTCTGCAAACTTTTGCCAGACCAACTCTTATTCAAA 905

QY 951 GAAAGGAGCCCAAGAATGTATAAGGTGTAA--AGACGACTCTCAATTTTCAGGATCCAG 1007

Db 906 TAAAGGAGAAACTTCTTTGCCACCAGTGTGACCCCTGACAAATACTCAGAGAAAGGATCTTC 965

QY 1008 TGAGTGTACAGAGCGGCCCTTCCCTGTACCACAAAAGACTATTTCCAGATCCATATCTCAT 1067

Db 966 TTCCTGTAACTGCGGCCAGCTTGACACAGACAAAGATTATTCTACACACACACAGGCGCTG 1025

QY 1068 TGATGAAGAAGGAAAGACACAGATAATGTACAAAGTGGATAGAGCCCCAAAATCTGCCGGGA 1127

Db 1026 CGATGCCAACGGAGAGACACAACACTCATGTACAAATGGGCCAAGCCGAAAATCTGTAGCGA 1085

QY 1128 GGATCTCACAGATGCTATTAGATTGCCCTTCCCTCTCTGGAGAGAAAGAGATTGTCCGCTTG 1187

Db 1086 GGACCTTGAGGGGCGAGTGAAGCTGCCTGCCTCTGCTGTGAGAGACCCACTGCCACCCCTG 1145

QY 1188 CAACCCCTGGATTTTATAACAAATGGATCATCTTCTTGCCATCCCTGTCTCTGGAAACATTT 1247

Db 1146 CAACCCAGGCTTCTTCAAAACCAACAACAGCACCTGCCAGCCCTGCCCATATGTCCTCCCTA 1205

QY 1248 TTCAGATGGAACCAAGAATGTAGACCATGTCCAGCAGGGAACGGAGCCCTGCACCTTGGCTT 1307

Db 1206 CTCCATGG--CTCAGACTGTACCCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGATTT 1262

QY 1308 TGAATATAAATGGTGGAAATGTCTTCTTCCCTGGCAACATGAAAACCTTCCCTGCTCAATGTTGG 1367

Db 1263 TGAATACAAATGGTGGAAACACCGTGGCCCAACAACATGGAACAGACCCGTTCTCAGTGGGAT 1322

QY 1368 GAATTCAAAGTCCGATGGAAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427

Db 1323 CAACCTCGAGTACAAGGGCATGACAGGCTGGGAGGTGGCTGGTGGTGGTGGTGGTGGTGGTGG 1382

QY 1428 GGCTGGAGGTCTTGACAAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACC 1487

Db 1383 TGCTGGAGCCTCAGACAATGACTTCATGATTCTCAGTCTGGTTGTGCCAGGATTTAGACC 1442

QY 1488 ACCAACATCTATGACTGGAGCCAC--GGGTTCTGAACTAGGAAGAATAACATTTGTCTTT 1544





Db 214 GCTTCATGCCGTGCAAGAGTCTGAGTACCACCTATGAGTACACGGCGTGTGACAGCACGGG 273

QY 231 CTCCAGGTGGAGAGTTGCCATTCCAAATTTCTGCAGTGGACTGCTCTGGCCCTGCCTGACCC 290

Db 274 TTCCAGGTGGAGGTCGCCGTGCCGCATACCCCGGGCCTGTGCACCAGCCTGCCCTGACCC 333

QY 291 AGTGAGAGGCAAGAATGCACCTTTCTCCTGTGCTTCTGAGAGATATCTAGAAATGAAGAA 350

Db 334 CGTCAAGGGCACCGAGTGTCTCTCTGCAACGGCGGGGAGTTTCTGGATATGAAGGA 393

QY 351 CCAGGTATGCAGTAAGTGTGGTGAAGGCACCTATTCTAACATCGCAACATTTCAATGGACACTGTGGT 410

Db 394 CCAATCATGTAAGCCATGCGCTGAGGGCCGCTACTCCCTCGGCACAGGCAATTCGGTTTGA 453

QY 411 TGAATGGGATGAATTGCCGGCAGGATTTTCTAACATCGCAACATTTCAATGGACACTGTGGT 470

Db 454 TGAGTGGGATGAGCTGCCCATGGCTTTGCCAGCCTCTCAGCCAAACATGGAGCTGGATGA 513

QY 471 GGGCCCTTCTGACAGCAGGCCAGACGGCTGTAAACAACTCTTCTTGGATCCCTCGTGGAAA 530

Db 514 CAGTGTCTGCTGAGTCCA--CCGGGAAGTGTACTTCGTCCAAGTGGTTTCCCGGGGGCA 570

QY 531 CTACATAGAATCTAATCGTGATGACTGCACGGTGTCTTGTGATCTATGCTGTGCACCTTAA 590

Db 571 CTACATCGCCTCCAACACGGAGCAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 630

QY 591 GAAGTCAGGCTATGTCTTCTTTGAGTACCAGTATGTCGACAACAACATCTTCTTTGAGTT 650

Db 631 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 690

QY 651 CTTTATTCAAAATGATCAGTGCAGGAGATGGACACCACTGACAAAGTGGGTAAACT 710

Db 691 TTTCTGTTCAGAAATGACCAGTGCACGCCCAATGCAGATGACTC---CAGTGGATGAAGAC 747

QY 711 TACAGAAATGGAGAATGGGGCTCTCATTTCTGTAATGCTGAATCAGGCACAAACATFACT 770

Db 748 CACAGAGAA--AGGATGGGAATTCACAGTGTGGAGCTAAATCGAGGCAATAATGTGCT 804

QY 771 CTACTGGAGAACTACAGGCATCCTTATGGGTTCTAAGGCGGTCAAGCCTGTGCTGGTAAA 830

Db 805 CTATTGGAGAACACAGCCTTCTCAGTATGGACCAAGTACCAAGCCTGTGCTGGTGAG 864

QY 831 AAATATCACAAATGAAGGGGTGGCGTACACATCAGAAATGTTTCTTGTCAAGCCAGGCAC 890

Db 865 AAACATTGCCATAACAGAGAA----- 885

QY 891 ATTCAGCAACAAACCAGGTTCAATTCAACTGCCAGGTGTGTCCAGAAACACCTATTCTGA 950

Db 886 ----- 885

QY 951 GAAAGGAGCCAAAGAATGTATAAGGTGTAAAGACGACTCTCAATTTTCAGGATCCAGTGA 1010

Db 886 -----AGGATCTTCTTC 897

QY 1011 GTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATACTCCATGTGA 1070

Db 898 CTGTAACGTGCGCCCGAGCTTGCACAGACAAAGATTTATTTACACACACACAGCGCCTGCGA 957

QY 1071 TGAAGAAAGGAAGACACAGATAATGTACAAAGTGGATAGAGCCCAAAATCTGCCGGGAGGA 1130

Db 958 TGCCCAACGGAGAGACACAACACTCATGTACAAATGGGCCAAGCCGAAATCTGTAGCGGAGGA 1017

QY 1131 TCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAAGAGATTTGCCGCCCTTGCAA 1190

Db 1018 CCTTGAGGGGCGAGTGAAGCTGCCCTGCTGTGGTGTGAAGACCCACTGCCACCCCTGCAA 1077

QY 1191 CCCCTGGATTTTATAACAAATGGATCATCTTCTTGCCATCCCTGTCTCTCTGGAACATTTTC 1250

Db 1078 CCCAGGCTTCTTCAAAACCAACACAGCACCTGCCAGCCCTGCCCATATGTTCTCTACTC 1137

QY 1251 AGATGGAAACCAAGAATGTAGACCATTGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTGA 1310

Db 1138 CAATGG---CTCAGACTGTACCCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGATTGA 1194

QY 1311 ATATAAATGGTGGAAATGTCTCTTCCCTGGCAACATGAANAACCTTCCCTGCTTCAATGTTGGGAA 1370

Db 1195 ATACAAATGGTGGAAACACGCTGCCACAAACATGGAACGACCGGTTCTCAGTGGGATCAA 1254

QY 1371 TTCAAAGTGCATGGAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGC 1430

Db 1255 CTTGAGTACAAGGGCATGACAGGCTGGGAGGTGGCTGGTGTATCAGTGTACAGCTGC 1314

QY 1431 TGGAGGTTCTGACAATGATTAACCTTAACTTGCATATCCCAGGATTTAAACCAACC 1490

Db 1315 TGGAGCCTCAGACAATGACTTCATGATTTCTACCTCTGGTGTGCCAGGATTTAGACCTCC 1374

QY 1491 AACATCTATGACTGGAGCCAC--GGGTCTTGAACTAGGAAGNAATAACATTTTGTCTTTGA 1547

Db 1375 GCAGTCGGTGATGGCAGACACAGAGAATAAAGAGGTGCCCAGAATCACATTTGTCTTTGA 1434

QY 1548 GACCCCTCTGTTCAGCTGACTGTGTTTTGTACTTTCTCATGCTGGTGGATATTAATAGAAAAGTAC 1607

Db 1435 GACCCCTCTGTTCTGTGAACCTGTGAGCTCTACTTCTATGCTGGTGGTGTGAATTTCTAGGACCAA 1494

QY 1608 AAATGTGGTAGAATCGTGGGTGGAAACCAAGAAAGAAACAAAGCTTACACCCCATATCATCTTT 1667

Db 1495 CACTCCTGTGGAGACGTGGAAAGGTTCCAAAGGCAACAGTCTTATACCTACATCATTTGA 1554

QY 1668 CAAGAATGCAACTTTTACATTTACATGGGCAATTCAGAGAACTAATCAGGGTCAAGATAA 1727

Db 1555 GGAGAACACTACCACGAGCTTCACCTGGGCTTCCAGAGGACCCACTTTTCTATGAGGCAAG 1614

QY 1728 TAGACGGTTTCATCAATGACATGGTGAAGATTTATTTCTATCAGGCCACTAATGCAGTTGA 1787

Db 1615 CAGGAAGTACACCAATGACGTGCCAAAGATCTACTCCATCAGTGTACCACATGTTATGAA 1674

QY 1788 TGGGTGGCGTCCCTCATGCCGTGCCCTGTGCCCTCGGTTCTGAAACAGTGGGTTCATCGTG 1847

Db 1675 TGGGTGGCCCTCCTACTGCCGTCCCTGTGCCCTTAGAAGCCTCTGATGTGGGCTCCTCCCTG 1734

QY 1848 TGTCCCTTCCCTCCAGGCCACTACATTTGAGAAAGAAACCAACCAAGTGCAGGAAATGTCC 1907

Db 1735 CACTCTTGTCTCTGCTGGTTACTATATTTGACCGAGATTTACAGGAACCTGCCACTCCTGCC 1794

QY 1908 ACCTGACACCTACCTGTCCATACATCAGGCTCTATGGCAAGAGAGGCTTGTATTTCCATGCGG 1967

Db 1795 CCCTAACACAAATTTCTGAAAGCCCAACAGCCTTATGGTGTCCAGGCTGTGTGCCCTGTGG 1854

QY 1968 GCCTGGGAGTAAAAACAATCAGGACCAATTCGGTTTGTCTATAGTACTGCTTTTCTTACCA 2027

Db 1855 TCCAGGGACCAAGAACAAAGATCCACTCTCTGTGCTACAAATGATGCAACCTTCTCAGC 1914

QY 2028 TGAAGAAGAAATCAGATTTTGCACATGACTTTTAGCAACCTCAGCAGTGTGGGCTCATTT 2087

Db 1915 CAACACTCCAACCCAGGACTTTCAACTACAACTTCTCCGCTTTGGCAAAACACCGTTCACIT 1974

QY 2088 AATGAATGGCCCCCAGCTTCACCTCCAAGGAACAAATACTTCCATTTCTTCAATATCAG 2147

Db 1975 TGCTGGAGGGCCCAAGCTTCACTTCCAAGGGTTGAAATACTTCCATCACTTTACCCCTCAG 2034

QY 2148 TTTTATGTGGGCATGAGGGGAAGAGATGGCTCTCTGTACCAACAATATTAACAGACITTTAC 2207

Db 2035 TCTCTGTGGAACCCAGGGTAGGAAATGTCTGTGTGCACCCGACAAATGTCACTGACCTCCG 2094

QY 2208 AGTAAAGAAATAGTGGCAGGGTCAAGATGNTTACACAAAATTTGGTAGGGGCAATTTGTATG 2267

Db 2095 GATTCTCTGAGGGTGAGTCAGGG-----TTCTCCAAATCTATCAGAGCCTACGTCG 2145

QY 2268 CCAGTCAACAATTTATCTCTTGAAGTAAGGGTTTCCGAGCAGCCTTATCATCACAAATC 2327

Db 2146 CCAGGCAGTCATCATCCCCCAGAGGTGACAGGCTACAAAGGCCGGGGTTTCTCTCACAGCC 2205

QY 2328 CATCATCTGGCAGATACATTCATAGGAGTCACAGTTGAAACCAACATTTGAAAAATATTTAA 2387

Db 2206 TGTACGCTTGTGATCGACTTATTTGGGGTGACAAACAGATATGACTCTGGATGGAAATCAC 2265



QY	2388	TATAAAGAAGATATGTTCCAGTTCCAAACAGCCAAATACCAGATGTGCATTTCTTTTA	2447
Db	2266	CTCCCCAGCTGAACCTTTCCACCTGGAGTCCTTGGGAATACCCGGACGTGATCTTCTTTTA	2325
QY	2448	TAAGTCTTCTACAGCAACAACATCTTGTATTAATGGCCCGATCAACTGCTGTGAAAAATGAG	2507
Db	2326	TAGGTCCAATGATGTGACCCAGTCTCTGCAGTCTTGGGAGATCAACCACCATCCCGCGTCAG	2385
QY	2508	GTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTCACTCCCCAGCAAGTGCCCGCAGCAGG	2567
Db	2386	GTGCAGTCCACAGAAAACCTGTCCCTGGAAAGTTTGCTGCTGCCAGGAACGTGCTCAGATGG	2445
QY	2568	TACCTGTGATGGGTGACGTTCTATTTCCCTGTGGGAGAGTGCTGAAGCTTGCCCTCTGTG	2627
Db	2446	GACCTGTGATGGCTGCAACTTCCACTTCCCTGTGGGAGAGCGCGGCTGCTTGCCCGCTCTG	2505
QY	2628	TACGGAGCATGACTTCCATGAGATTGAGGGAGCCCTGCAAGAGAGAGGATTTTCAGGAACCTT	2687
Db	2506	CTCAGTGGCTGACTACCATGCTATCGTCAGCAGCAGCTGTGTGGCTGGGATCCAGAAGACTAC	2565
QY	2688	GTATGTGTGGAATGAACCTAAATGGTGCATTAAAGGAATTTCTTTGCCCTGAGAAAAAGTT	2747
Db	2566	TTACGTGTGGCGAGAACCCCAAGCTATGCTCTGGTGGCATTTCTCTGCCCTGAGCAGAGAGT	2625
QY	2748	GGCAACCTGTGAAACGGTTGAC'TTTTGGCTGAAGGTGGGAGCCGGTGTGGAGCTTTTAC	2807
Db	2626	CACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTGTAC	2685
QY	2808	TGCCGTTTGTGCTGGTGCTCTGACCTGCTACITCTGGA AAAAAGAATCAAAAACCTGGAATA	2867
Db	2686	TGCCATCTGCTCACCGCTCTTGACCTGCTACTTTTGGAAAAAGAATCAAAAACCTAGAGTA	2745
QY	2868	CAATATTCCAAAGTTAGTAATGACGACTAACTCAAAAGAGTGTGAAC T 2915	
Db	2746	CAAGTACTCCAAGCTGGTGATGAATGCTACTCTCAAGGATTGTGGCCT 2793	
RESULT 10			
BC029010			
LOCUS			
DEFINITION Mus musculus, clone IMAGE:5361249, mRNA, partial cds.			
ACCESSION BC029010			
VERSION BC029010.1 GI:20810059			
KEYWORDS house mouse.			
SOURCE			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE 1 (bases 1 to 2188)			
AUTHORS Strausberg, R.			
TITLE Direct Submission			
JOURNAL Submitted (01-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
REMARK			
COMMENT NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a> Contact: MGC help desk Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a> Tissue Procurement: The Cepko Laboratory cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: <a href="http://www.hgsc.bcm.tmc.edu/cdna/">http://www.hgsc.bcm.tmc.edu/cdna/</a> Contact: <a href="mailto:amg@bcm.tmc.edu">amg@bcm.tmc.edu</a> Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.			
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAK.Plate: 54 Row: O Column: 14			

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.			
FEATURES		Location/Qualifiers	
source		1..2188	
		/organism="Mus musculus"	
		/db_xref="taxon:10090"	
		/clone="IMAGE:5361249"	
		/tissue_type="Eye, retina, mouse strain C57Bl\6"	
		/clone_lib="NIH_MGC_94"	
		/lab_host="DH10B"	
		/note="Vector: pCMV-SPORT6"	
CDS		<1..720	
		/codon_start=1	
		/product="Unknown (protein for IMAGE:5361249)"	
		/protein_id="AAH29010.1"	
		/db_xref="GI:20810060"	
		/translation="DTALQNVNIKEDMFPVSPQPDVHFFYKSSATTSCINGRSTAVKMRGNPMRPGAGVISVPSKCPAGTCDCGTFYFLWESAECPLCKEHDFHEIEKKKRGLOEILYVWNEPKWCIKIGISLPEKKLSCTCETVDFWLKVGAGVGAFTAVLLVLTCTYFWKKNOKLEYKYSKLVMTTNSKECELPAAADSCAIMEGEDNEEDVVYVSNKQSLGLKLSLATKEKDDHFEVQLKSSRCPN1"	
BASE COUNT		695 a	360 c 453 g 680 t
ORIGIN			
Query Match 20.0%; Score 711.6; DB 10; Length 2188;			
Best Local Similarity 80.8%; Pred. No. 3.8e-156;			
Matches 921; Conservative 0; Mismatches 174; Indels 45; Gaps 6;			
QY	2365	GA AACACACATTTGAAAAATATTAATATAAAAGAAGATATGTTCCAGTTCCAAAGCCAA	2424
Db	1	GACACTGCCTTGCAAAACGTTAATATCAAGGAGGATATGTTCCAGTGTCAACGAGCCAA	60
QY	2425	ATAC CAGATGTGCATTTCTTTATAAGTCTTCTACAGCAACAACATCTTGTATTAAATGGC	2484
Db	61	GTACCAGATGTTCAATTTCTTTATAAGTCTTCCACAGCCACCACATCTTGTATTAAATGA	120
QY	2485	CGATCAACTGCTGTGAAAATGAGGTGTAATCCTACTATAATCTGGAGCAGGAGTGATTTCA	2544
Db	121	CGATCGACTGCTGTGAAGATGAGGTGTAATCCCATGAGACCTGGTG CAGGTGTGATATCA	180
QY	2545	GTCCCCAGCAAGTGCC CAGCAGGTACCTGTGATGGTGTGTACGTTCTATTTCCTGTGGGAG	2604
Db	181	GTCCCCAGCAAGTGTCCAGCTGGCACCTGTGATGGTGTGTACTTCTACTTCTATGGGAG	240
QY	2605	AGTGCTGAAGCTTGCCCTCTGTGTACGGAGCATGAC TTCATGAGATTGAGGGAGCCTGC	2664
Db	241	AGTG CAGAAGCTTGCCCTCTGTGCAAGAACATGAC TTCATGAGATCGAGAAAAA AAAA	300
QY	2665	AAGAGAGGATTT CAGGAAACCTTGATGTGTGGAATGAACCTAAATGGTGCATTAAGGA	2724
Db	301	AAAAGAGGGCTTCAGGAAATATTATATGTATGGAATGAACCTAAGTGGTGCATTAAGGA	360
QY	2725	ATTCTTTGCCCTGAGAAAAAGTTGGCAACCTGTGAACCGGTGACTTTTGGCTGAAGGTG	2784
Db	361	ATTCTTTGCCCGAAAGAGAGTTGTCAACCTGCGAAACTGTTGACTTTTGGCTGAAAGTG	420
QY	2785	GGAGCCGGTGTGGGAGCTTTTACTGCCGTTTGTGCTGGTGGCTCTGACCTGCTACTTCTGG	2844
Db	421	GGAGCTGGTGTGGGCGCGTTCACAGCCGTTTGTGTTGGTGGCTTTAACATGCTACTTTTG	480
QY	2845	AAAAAGAATCAAAAACTGGAATACAAATATTCCAAAGTTAGTAATGACGACTAACTCAAAA	2904
Db	481	AAAAAGAATCAAAAGCTGGAATACAAATATTCTAAATTAGTAATGACGACTAACTCAAAA	540
QY	2905	GAGTGTGAACCTCCCGGCTGCAGACAGTTGTGCTATCATGGAAGGAGAGATAATGAAGAG	2964
Db	541	GAGTGTGAACCTCCCGGCTGCAGACAGCTGTGCTATCATGGAAGGAGAGATAATGAAGAG	600
QY	2965	GAAGTTGTATATTTCCCAATAAACAGTCACTACTAGGAAAACTCAAAATCTTTGGCAACCAAG	3024
Db	601	GATGTTGTATATTTCCCAACAACACAGTCACTACTAGGAAAACTTAAGTCTTTGGCCACAAAG	660

QY	3025	GAAAAAGAACCACTTTTGAATCTGTTCAACTGAAACACCTCAAGATCCCCAAATATATGA	3084
Db	661	GAAAAAGATGACCACTTTTGAGTCTGTTCAAGTGAATCTCAAGATGCCCAATATATGA	720
QY	3085	AGAGACAGTGCTGTAGCCTTTGAGACTAATGAACAAAGAAACCTGCTCTAGTTTACAGGA	3144
Db	721	AGAGACAATGCTGTAGCAATCA-----GG	744
QY	3145	CCATATTTTAGGGTCTGTCCTCATACCTGTGCACATTTGGTGATCTCACAGAGGAGGCCAT	3204
Db	745	ACCCCATTAGGACTTGTCTCTACATCTGACATGGCTGTGATCTCACAAAGGAGG--TGT	802
QY	3205	GCCGCTGAAAAGGGAAGGAGATTGAAACATTTGATTGCCTTATCACATGGTCAAGTACCT	3264
Db	803	GCCAAATGAGAGGAGGAGG---CAAAAATGCTGATTGCCTTATAACAT-GTCACATACTT	857
QY	3265	TGCCAAATAAGGAAAGCAAAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGA	3324
Db	858	TGTCAAACTACAGAAGACAGATGATGTGGGTCTC-ACTGAAGATGAAGCTTAACCTCAGGA	916
QY	3325	AGAGATTTATCTGTATATACACATAACTGAAACCCAAAGTTTAAGCCCCACCAATGCAC	3384
Db	917	AGAGATTTAGCTGTATATACACATAACTGAAACCCCAAGTTTAAGCCCCACCAATGCAC	976
QY	3385	TGATGCATGCCATATATAATGGGTAACCTTTTATCTTTATGATGTCTACATAACAAGT	3444
Db	977	TGA-GCATGCCATATGATTACTGGGTAACCTTTAATTTCTTTATACTGTCTACATACAAAAT	1035
QY	3445	GTGATTTGGAAGGCACATGTGAGCATATGCATTATGATCCAAATTTATGTTTCTTGT	3504
Db	1036	ATGTGTTGGAAGGCAGACATAAAACATATGCCTTGTGTTCACAGTATGTCTTGTCTGT	1095
RESULT 11			
BC031648			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REMARK			
COMMENT			

FEATURES		Location/Qualifiers
source		1..2449
		/organism="Homo sapiens"
		/db_xref="LocusID:57535"
		/db_xref="taxon:9606"
		/clone="MGC:35166 IMAGE:5169952"
		/tissue_type="Brain, adult medulla"
		/clone_lib="NIH_MGC_119"
		/lab_host="DH10B"
		/note="vector: pCMV-SPORT6"
CDS		208..2199
		/codon_start=1
		/product="KIAA1324 protein"
		/protein_id="AAH31648.1"
		/db_xref="GI:21594948"
		/translation="MYKWAKPKICSFDLEGAVKLPASGVKTHCPKPCNPGFFKTNNSTC
		QPCPYGSYNSGSDCTRCPPAGTEPAVGFEYKWWNTLPTNMETTVLSGINFEYKGMTGWE
		VAGDHIYTAAGASDNDFMILTLVPGFRPPQSVMTADTENKEVARITFVFETLCSVNC
		LYFMVGVNSRTNTPVETWKGSKGOSYTYIIEENTTTSTFWAQTTFHEASRKYTND
		VAKIYSINVTNVMNGVASYCRPCALEASDVSSCTSCPAGYYIDRDSGTCHSCPNTI
		LKAHQPYQVQACVPCPGPTKNNKIHSICYNDCITFSRNTPTFRFNYNFSALANTVTLAG
		GPSETSKGLYFHHFTLSLCGNQGRKMSVCTDNVTLRIPEGESGFSKSTAYVCOAV
		IIPPEVTGYKAGVSSQPVSLADRLTGVTDTMDTLDGITSPAELFHLFSLGIPDVIFFYR
		SNVTQSCSSSRSTTIRVRCSPQKTVPGSLLLPDGTCDGCFNFIHFIWESAAACPL
		CSVADYHAIVSSCVAGIQKTTYVWREPKLCSGGISLPEQRVTTCKTIDFWLKVGISAG
		TCTAILLLTVLTCYFWKKNQKLEYKSLVMNATLKDCLDLPADSCAIMEGEDVEDDLI
		FTSKKSLFGKIKSETSKRTPDGFSDVPLKTSGGPMDL"
BASE COUNT		618 a 690 c 563 g 578 t
ORIGIN		
Query Match		19.3%; Score 687; DB 9; Length 2449;
Best Local Similarity		60.2%; pred. No. 2.4e-150;
Matches 1197; Conservative		0; Mismatches 775; Indels 15; Gaps 3;

QY	1082	AGACACAGATAATGTACAAAGTGGATAGAGCCCCAAAATCTGCCGGAGGATCTCACAGATG	1141
Db	197	AGACACAACATCATGTACAAATGGCCCAAGCCGAAAATCTGTAGCGAGGACCTTGAGGGGG	256
QY	1142	CTATTAGATTGCCCCCTTCTGGAGAGAAGAAGGATGTCCGCCCTTGCAACCCCTGGATTTT	1201
Db	257	CAGTGAAGCTGCCCTGCTGGTGTGAAGACCCACTGCCCAACCCTGCAACCCAGGCTTCT	316
QY	1202	ATAACAANTGGATCATGTTTGTCCATCCCTGTCTCTCTGGAACATTTTCAGATGGAACCA	1261
Db	317	TCAAAACCAACAACACAGCACCTGCCAGCCCTGCCCATATGGTTCCTACTCCAATGG---CT	373
QY	1262	AAGAATGTAGACCATGTCCAGCAGGAACGGAGCTGCACCTTGGCTTTGAATATAAATGGT	1321
Db	374	CAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGATTTGAATACAAATGGT	433
QY	1322	GGAATGTCTTCTCCCTGGCAACATGAAAACCTTCCCTGCTCAATGTGGGAAATTCAAAGTGG	1381
Db	434	GGAACACGCTGCCCAACAACATGGAACGACCCGTCTTCAGTGGGATCAACTTCGAGGTACA	493
QY	1382	ATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTG	1441
Db	494	AGGCGATGACAGGCTGGGAGGTGGCTGGTGATCACATTTACACAGCTGCTGGAGCCCTCAG	553
QY	1442	ACAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCACCACAATCTATGA	1501
Db	554	ACAATGACTTCATGATTCCTCACTCTGGTTGTGGCAGGATTTAGACCTCCCGAGTCGGTGA	613
QY	1502	CTGGAGCCAC---GGGTTCTGAACCTAGGAAGATAAACATTTGTCTTTGAGACCCCTCTGTT	1558
Db	614	TGGCAGACACAGAGAATAAAGAGGTGGCCAGAATCACATTTGTCTTTGAGACCCCTCTGTT	673
QY	1559	CAGCTGACTGTGTTTGTACTTTCATGGTGGGATPATTAATAGAAAAAGTACAAATGTGGTAG	1618
Db	674	CTGTGAACCTGTGAGCTCTACTTTCATGGTGGGTGTGAANTCTAGGACCAACACTCCTGTGG	733
QY	1619	ANTCGTGGGTGGAAACCAAGAAAACAAGCTTACACCCCATATCATCTTCAACAANTGCAA	1678
Db	734	AGACGTGGAAGGTTCCAAAGGCAACACAGTCTCTATACCTACATCATTTGAGGAGAACACTA	793

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 51 Row: d Column: 23  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

QY 1679 CTTTACATTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTCA 1738  
Db 794 CCACGAGCTTCACCTGGCCCTCCAGAGGACCACCTTTTCATGAGGCAAGCAGGAAGTACA 853  
QY 1739 TCAATGACATGGTGAAGATTTATTCTATCATCAGCCCACTAATGCGAGTTGATGGGGTGGCGT 1798  
Db 854 CCAATGACGTGCCAAGATCTACTCCATCAATGTCACCAATGTTATGAATGGCGTGGCCT 913  
QY 1799 CCTCATGCGGTGCGCTGTGCCCTCGGTCTCTGAACAGTCGCGGTTCATCGTGTGCCCTGCC 1858  
Db 914 CCTACTGCGGTGCGCTGTGCCCTAGAACCTCTGATGTGGCTCCCTGCAACCTCTTTGTC 973  
QY 1859 CTCAGGCCCACTACATTTGAGAAAGAAACCAACCAAGTCAAGGAAATGTCACCTGACACCT 1918  
Db 974 CTGCTGGTTACTATATTGACCGAGATTCAGGAACCTGCCACTCTCGCCCTTAACACAA 1033  
QY 1919 ACCTGTCCATACATCAGGTCTATGGCAAGAGGCTTTGTATTCCATGCGGGCCTGGAGTA 1978  
Db 1034 TTCTGAAAGGCCACCAAGCCTTATGGTGTCCAGGCTGTGTGCCCTGTGGTCCAGGACCA 1093  
QY 1979 AAAACAATCAGGACCATTCGGTTTGTATAGTACTGCTTTTTCATACCATGAAAGAAA 2038  
Db 1094 AGAACACAAGATCCACTCTCTGTGCTACAATGATTGCACCTTCTCAGCAACACTCCAA 1153  
QY 2039 ATCAGATTTTGCACTATGACTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGGCC 2098  
Db 1154 CCAGGACITTCAACTACAACCTTCTCCGCTTTGGCAACACCCGTCACTCTTGCTGGAGGC 1213  
QY 2099 CCAGCTTCACTCCAAAGGAACAATACTTCCATTCTTCAATATCAAGTTTATGTGGGC 2158  
Db 1214 CAAGCTTCACTTCCAAAGGTTGAATACTTCCATCACTTTACCCTCAGTCTCTGTGGAA 1273  
QY 2159 ATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATAATACAGACTTTACAGTAAAGAAA 2218  
Db 1274 ACCAGGTAGGAAAATGTCTGTGTGCACCGACAATGTCACTGACCTCCGGATTCTCTGAGG 1333  
QY 2219 TAGTGGCAGGTCAGATGATTACACAAATTTGGTAGGGGCATTTGTATGCCAGTCAACAA 2278  
Db 1334 GTGAGTCAGG-----TTCTCCAAATCTATCACAGCCTACGTGTGCCAGGCAGTCA 1384  
QY 2279 TTATTCTTCTGAAAGTAAGGTTTCCGAGCAGCCTTATCATCACAATCCATCATCTCTGG 2338  
Db 1385 TCATCCCCCAGAGGTGACAGGCTACAGGCGGGTTTCTCTCACAGCCTGTGAGCCTTG 1444  
QY 2339 CAGATACATTCATAGGAGTCACAGTTGAAACCACATTTGAAATAATTAATATAAAGAA 2398  
Db 1445 CTGATCGACTTATTTGGGTGACAACAGATATGACTCTGGATGGATCACCTCCCCAGCTG 1504  
QY 2399 ATATGTTCCAGTTCACAACAGCCCAATACCAGATGTGCATTTCTTTTATAAGTCTCTA 2458  
Db 1505 AACTTTTCCACCTGGAGTCCCTTGGGAATACCGGACGTGATCTTCTTTATAGGTCCAATG 1564  
QY 2459 CAGCAACACATCTTGTATTAAATGGCCGATCAACTGCTGTGAAATGAGGTGTAATCCTA 2518  
Db 1565 ATGTGACCCAGTCCCTGCAGTTCTGGGAGATCAACCACCATCCGCGTCAGGTGCAGTCCAC 1624  
QY 2519 CTAATCTGGAGCAGGAGTGATTTTCAGTCCCCCAGCAAGTGCCCCAGGTAACCTGTGATG 2578  
Db 1625 AGAAAACCTGCCCTGGAAGTTTGTCTGCCAGGAACGTGCTCAGATGGGACCTGTGATG 1684  
QY 2579 GGTGTACGTTCTATTCTCTGTGGGAGAGTGTGAAGCTTGCCCTCTGTGTACGGAGCATG 2638  
Db 1685 GCTGCAACTTCCACTTCTCTGTGGGAGAGCGGGCTGCTTGCCCGCTCTGCTCAGTGGCTG 1744  
QY 2639 ACTTCCATGAGATTTGAGGGAGCCTGCAAGAGAGGATTTTCAGGAACCTTGTATGTGGA 2698  
Db 1745 ACTACCATGCTATCGTCAGCAGCTGTGTGGCTGGGATCCAGAAAGACTACTTACGTGTGGC 1804  
QY 2699 ATGAACCTAAATGGTGCATTAAGGAATTTCTTTGCCCTGAGAAAAGTTGGCAACCTGTG 2758  
Db 1805 GAGAACCAAGCTATGCTCTGGTGGCATTTCTCTGCCTGAGCAGAGAGTCAACCATCTGCA 1864

QY 2759 AAACGGTTGACTTTTGGCTGAAGGTGGAGCCGGTGTGGAGCTTTTACTGCCGTTTTC 2818  
Db 1865 AAACCATAGATTTTGGCTGAAGGTGGGCATCTCTGCAGGCACCTGTACTGCCATCTGC 1924  
QY 2819 TGGTGGCTCTGACCTGCTACTTCTCTGGAAAAGAATCAAAAACCTGGAATACAAATATTCCA 2878  
Db 1925 TCACCGTCTTGACCTGCTACTTTTGGAAAAGAATCAAAAACCTAGAGTACAAAGTACTCCA 1984  
QY 2879 AGTTAGTAATGACGACTAACTCAAAGAGTGTGAACCTCCCGGCTGCAGACAGTGTGCTA 2938  
Db 1985 AGCTGGTGATGAATGCTACTCTCAAGGACTGTGACCTGCCAGCAGCTGACAGCTGCGCCA 2044  
QY 2939 TCATGGAAGGAGAAGATAATGAAGAGGAAGTTGTATATTTCCAATAAACAGTCACTACTAG 2998  
Db 2045 TCATGGAAGGCGAGGATGTAGAGGACGACCTCATCTTTACCAGCAAGAAGTCACTCTTTG 2104  
QY 2999 GAAAACCTCAAATCTTTGGCAACCAAGGAAAAGAAGACCATTTTGAATCTGTCAACTGA 3058  
Db 2105 GGAAGATCAAATCATTTACCTCCAAGAGGACTCCTGATGGATTGACTAGTGCAGCTGA 2164  
QY 3059 AAACCTC 3065  
Db 2165 AGACATC 2171  
RESULT 12  
AK026832  
LOCUS AK026832 2129 bp mRNA linear PRI 29-SEP-2000  
DEFINITION Homo sapiens cDNA: FLJ23179 fis, clone LNG10890.  
ACCESSION AK026832  
VERSION AK026832.1 GI:10439779  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens human lung cDNA to mRNA, clone\_lib:LNG clone:LNG10890.  
ORGANISM Homo sapiens  
REFERENCE 1 (sites)  
AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,  
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Tanigami,A.,  
Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hirao,M.,  
Ohmori,Y., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T.,  
Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2129)  
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,  
Shibahara,T., Tanaka,T. and Nakamura,Y.  
TITLE Direct Submission  
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
Japan (E-mail:cdnal@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing: Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing: Departent of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).  
FEATURES  
source Location/Qualifiers  
1.2129  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="LNG10890"  
/tissue\_type="human lung"  
/clone\_lib="LNG"  
/note="cloning vector pME18SFL3"  
BASE COUNT 534 a 593 c 491 g 511 t  
ORIGIN  
Query Match 17.8%; Score 633.8; DB 9; Length 2129;  
Best Local Similarity 60.1%; Pred. No. 7.3e-138;



	Matches 1112; Conservative	0;	Mismatches	722;	Indels	15;	Gaps	3;
QY	1220	CTTGCCATCCTGTCTCTCTGGAACAATTTCAGATGGAAACCACCAAGAATGTAGACCATGTC	1279					
Dd	33	CCTGCCAGCCTGCCCATATGGTTCTTACTCCAATGG--CTCAGACTGTACCGCTGCC	89					
QY	1280	CAGCAGGAACGAGGCCTGCACCTTGGCTTTGAATATAAATGGTGGATGTCTCTCCCTGGCA	1339					
Dd	90	CTGCAGGACTGAACCTGCTGTGGGATTGAATACAAATGGTGGAACACGCTGCCACAA	149					
QY	1340	ACATGAAAACCTCCTGCTTCAATGTTGGGAATTCAAAGTGGCATGGGAATGAATGGTTGGG	1399					
Dd	150	ACATGAAAACGACCGTTCTCAGTGGGATCAACTTCGAGTACAAGGCGATGACAGGCTGGG	209					
QY	1400	AGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGACAAATGATTAACCTGATCT	1459					
Dd	210	AGGTGGCTGGTATCACATTTACACAGCTGCTGGAGCCTCAGACAATGACTTTCATGATTC	269					
QY	1460	TAAACTTGCAATPCCCAGGATTTAAACCACCAACACATCTATGACTGGAGCCAC---	1516					
Dd	270	TCACTCTGGTTGTGCCAGGATTTAGACCTCCGCAGTCCGTGATGSGCAGACACAGAGAATA	329					
QY	1517	CTGAAC TAGGAATAAACATTTGTCTTTGAGACCCCTCTGTTTACGCTGACTGTGTTTTGT	1576					
Dd	330	AAGAGGTGGCCAGAAATCACATTTGTCTTTGAGACCCCTCTGTTCTGTAACGTGAGCTCT	389					
QY	1577	ACTTCATGGTGGATTAATAGAAAAAGTACAAAATGTGGTAGAATCGTGGGGTGGAAACCA	1636					
Dd	390	ACTTCATGGTGGTGTGAATTTCTAGGACCAACACTCCTGTGGAGACGTGGAAAGGTTCCA	449					
QY	1637	AAGAAAAACAAGCTTACACCCATATCATCTTCAAGAAATGCAAATTTTACATTTACATGGG	1696					
Dd	450	AAGGCAACAGTCCTATACCTACATCATATGAGGAGAACACTACCACGAGCTTCACTGGG	509					
QY	1697	CATTCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTTCATCAATGACATGGTGAAGA	1756					
Dd	510	CCTCCAGAGGACCACTTTTCNTAGGCAAGCAGGAAGTACACCAATGACGTTGCCAAGA	569					
QY	1757	TTTATTCTATCAGACCCTAATGCAATGATGGGTGGCGTCCCTCATGCGCTGGT	1816					
Dd	570	TCTACTCCATCAATGTACCAATGTTATGAATGGGTGGCGTCCCTACTGCGCTCCCTGTG	629					
QY	1817	CCCTCGGTTCTGAACAGTCGGGTTTCATCGTGTGTCCTCCCTGCCCCCAGGCCACTACATG	1876					
Dd	630	CCCTAGAAGCCTCTGATGTGGCTCCTCTGCACCTCTTGTCTCTGTTACTATATTG	689					
QY	1877	AGAAAGAAACCAACAGTCGCAAGGAATGTCACACCTGACACCTACCTGTCCATACATCAGG	1936					
Dd	690	ACCGAGATTCAAGAACCTGCCACTCCTGCCCCCTAACACAATTTGAAAGCCCAACAGC	749					
QY	1937	TCTATGGCAAGAGGCTTGTATTCATGGGGCCCTGGGAGTAAACAAATCAGGACCATT	1996					
Dd	750	CTTATGGTGTCCAGGCCCTGTGTGCCCTGTGGTCCAGGGACCAAGAACAAGATCCACT	809					
QY	1997	CGGTTTGCTATAGTGAAGTCTTTTCTACCATGAAAAAGAAATCAGATTTTGCACTATG	2056					
Dd	810	CTCTGTGCTACACGATTCACCTTCTCAGCAACACTCCGACCAAGGACTTTCACCTACA	869					
QY	2057	ACTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGGCCCCAGCTTCACCTCCAAAG	2116					
Dd	870	ACTTCTCCGCTTTGGCAACACTGTCACTCTGCTGGAGGGCCCAAGCTTCACCTCCAAAG	929					
QY	2117	GAACAAAATACCTCCATTTCTTCAATATCAGTTTATGTGGGCATGAGGGGAAGAAGATGG	2176					
Dd	930	GGCTGAAATACITCCATCATTACCTCAGTCTCTGTGGAAACCAGGGTAGGAAAATGT	989					
QY	2177	CTCTCTGTACCAACAATATAACAGACTTTACAGTAAAGAAATAGTGGCAGGGTCAGATG	2236					
Dd	990	CTGTGTGCACCGACAATGTCACTGACCTCCGGATTCTCTGAGGGTGTGAGTCAGGG	1042					
QY	2237	ATTACACAAATTTGGTAGGGGCATTTGTATGCCAGTCAACAATTTATCTCTGAAAGTA	2296					
Dd	1043	--TTCTCCAAATCTATCACAGCCTACGCTGTGCCAGGCAGTCAATCATATCCCCCAGAGGTGA	1100					



```
REFERENCE 2 (bases 1 to 5567)
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.
TITLE Direct Submission.
JOURNAL Submitted (31-JAN-2000) Osamu Ohara, Kazusa DNA Research Institute,
Laboratory of DNA Technology; 1532-3 Yana, Kisarazu, Chiba
292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,
URL:http://www.kazusa.or.jp/huge/, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)

FEATURES             Location/Qualifiers
     source            1..5567
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="fh14139"
     gene              /tissue_type="brain"
                        /clone_lib="pBluescriptII SK plus"
     CDS                1..5567
                        /gene="KIAA1324"
                        <1..1744
                        /gene="KIAA1324"
                        /note="Start codon is not identified."
                        /codon_start=2
                        /product="KIAA1324 protein"
                        /protein_id="BAA92562.1"
                        /db_xref="GI:7243029"
                        /translation="WNTLPTNNETTIVLSGINFEYKMGTMGEVAGDHIYTAAGASDND
FMILTLVVPGRPPQSVMDTENKEVARITEVFETLCSVNCLEYFMVGVNSRNTNPVE
TWKSGKQSYTYIIENNTTSFWAFQRTTFHEASRKYTNDAKIYSINVTNMGV
ASYCRPCALEASDVSSCTSCPAGYIIDRDSGTCHSCPNTILKAHQPYGVQACVPCG
PGTKNNKIHSLCYNDCFTSRNTPRTFNFSALANTVTLAGGSPFTSKGLKYFHHFT
LSLCNGOGRKMSVCTDNVTLRIPEGESGFSKSIYAVQAVIIPPEVTGYKAGVSSQ
PVSADRLLIGVTTDMTLDGITSPAELFHLESIGIPDVIFFEYRSNDVTQSCSSGRSTTI
RVRCSPQKTVPGSLLLPGTCDGTCDFGNFHLWESAAACPLCSVADYHAIIVSSCVAG
IQKTTYVWREPKLCSGGISLPEQRVTICKTIDFWLKVGISAGCTAAILTLVLTICYFWK
KNQKLEYKYSKLVMMNATLKDCLDLPAAADSCAIMEGEDVEDDLIFTSKKSLFGKIKSFTS
KQPAPVTIISLEDS"

BASE COUNT   1382 a   1404 c   1270 g   1511 t
ORIGIN
Query Match      16.7%;   Score 594;   DB 9;   Length 5567;
Best Local Similarity 60.0%;   Pred. No. 2e-128;
Matches 1031;   Conservative 0;   Mismatches 675;   Indels 12;   Gaps 2;

QY  1317 ATGGTGGAAATGTCCTTCCTGGCAACATGAAAACTTCCTGCTTCAATGTTGGGAATTCAAA 1376
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   1 ATGGTGGAAACACGCTGCCCAACACATGGAACGACCGTCTCAGTGGGATCAACTCGA 60

QY  1377 GTGGCATGGAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGG 1436
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   61 GTACAAGGCATGACAGGCTGGGAGGTGGCTGGTGATACATTTACACAGCTGCTGGAGC 120

QY  1437 TTCTGACAATGATTACCTGATCTTAACTTGCAATATCCAGGATTTAAACCAACCAATC 1496
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   121 CTCAGACAATGACTTCATGATCTCACTCTGGTTGGCCAGGATTTAGACCTCCGCAGTC 180

QY  1497 TATGACTGGAGCCAC---GGGTTCTGAACTAGGAAGATAACATTTGTCTTTGAGACCCCT 1553
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   181 GGTGATGGCAGACACAGAGAATAAGAGGTGGCCAGAAATCACATTTGTCTTTGAGACCCCT 240

QY  1554 CTGTTCACTGACTGTGTTTGTACTTTCATGGTGGATATTAATAGAAAAAGTACAAATGT 1613
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   241 CTGTTCTGTGAACCTGTGAGCTCTACTTCTCATGGTGGGTGTGAATTTCTAGGACCAACTCC 300

QY  1614 GGTAGAATCGTGGGTGGAAACCAAGAAACAAAGCTTACACCCCATATCATCTTCAAGAA 1673
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   301 TGTGGAGACGTGGAAAGGTTCCAAAGGCAACACAGTCCCTATACCTACATCATTTGAGGAGAA 360

QY  1674 TGCAACTTTTACATTTACATGGGCATTCACAGAGAACCTAATCAGGGTCAAGATAATAGACG 1733
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   361 CACTACCACGAGCTTCACCTGGGCCTTCACAGAGGACCACCTTTTCATGAGGCAACGAGGAA 420

QY  1734 GTTCATCAATGACATGGTGAAGATTTATTCTATATCAGCCCACTAATGCAGTTGAGGGGT 1793
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   421 GTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTACCAATGTTATGAATGGCGT 480
```

```
QY  1794 GCGTCCCTCATGCGCTGCTGTGCCCTCGGTTCTGAACAGTCGGGTTTCATCGTGTGTCCC 1853
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   481 GGCCTCCTACTGCCGTCCCTGTGCCCTAGAACCTCTGTATGTGGGCTCCTCCTGCACCTC 540

QY  1854 CTGCCCTCCAGGCCACTACATTGAGAGAAAGAAACCAACCAAGTGCAGGAATGTCCACCTGA 1913
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   541 TTGTCTCTGCTGGTTACTATATTGACCGAGATTTCAGGAACCTGCCACTCCTGCCCCCTAA 600

QY  1914 CACCTACCTGTCCATACATCAGGTCTATGGCAAGAGAGGCTTGATTCCATGCGGGCCCTGG 1973
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   601 CACAATTCTGAAAGCCCAACACAGCCTTATGGTGTCCAGGCCTGTGTGCCCTGTGGTCCAGG 660

QY  1974 GAGTAAAAACAATCAGGACCATTTCGGTTTGTATATAGTACTGCTTTTTCFACCATGAAAA 2033
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   661 GACCAAGAACCAACAGATCCACTCTCTGTGCTACAATGATTGCACCTTCTCAGCGCAACAC 720

QY  2034 AGAAAAATCAGATTTTGCACTATGACTTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAA 2093
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   721 TCCAACCCAGGACTTCAACTACAACCTTCTCCGCTTTGGCAACACACCGTCACTCTTGCTGG 780

QY  2094 TGGCCCCAGCTTCACTCCAAAGGAACAAAAATACTTCCATTCTTCAATATACAGTTATG 2153
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   781 AGGGCCAAGCTTCACTTCCAAAGGGTTGAATACTTCCATCACTTTACCCCTCAGTCTCTG 840

QY  2154 TGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATAATAACAGACTTTACAGTAAA 2213
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   841 TGGAAACCAAGGTAGGAAAATGTCTGTGTGCACCGACAATGTCACTGACCTCCGGATTCC 900

QY  2214 AGAAATAGTGGCAGGTCAAGTATGATTACACAAATTTGGTAGGGGCATTTGTATGCCAGTC 2273
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   901 TGAGGTGAGTCAGG-----TTCTCCAAATCTATCACAGCCTACGTCGTGCCAGGC 951

QY  2274 AACAAATTATTCCTTCTGAAAAGTAAGGGTTTCCGAGCAGCCTTATCATCACAAATCCATCAT 2333
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   952 AGTCATCATCCCCCAGAGGTGACAGGCTACAGGCGGGGTTTCCCTCACAGCCTGTTCAG 1011

QY  2334 TCTGGCAGATACATTCATAGGAGTCACAGTTGAAACCCACATTTGAAAAATATTAATATAA 2393
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   1012 CCTTGTGATCGACTTATTTGGGTGACAACAGATATGACTCTGGATGGAATCACCTCCCC 1071

QY  2394 AGAAGATATGTTCCAGTTCACCAACAGCCAAATACCAGATGTGCATTTCTTTTATAAGTC 2453
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   1072 AGCTGAACCTTTCCACCTGGAGTCCCTTGGGAATACCGGACGTGATCTTCTTTTATAGGTC 1131

QY  2454 TTCTACAGCAACAACATCTTGTATTAAATGGCCGATCAACTGCTGTGAAAAATGAGGTGTA 2513
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   1132 CAATGATGTGACCCAGTCTGCAGTTCTGGGAGATCAACCCACCATCCGCGTCAGTGCAG 1191

QY  2514 TCCTACTAAATCTGGAGCAGGAGTGATTTCACTCCCGAGCAAGTGCCCGAGCAGTACCTG 2573
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   1192 TCCACAGAAAACTGTCCCTGGAAAGTTTGTGCTGCCAGGAACGTGCTCAGATGGGACCTG 1251

QY  2574 TGATGGGTGTACGTTCTATTCTTCTGTGGGAGAGTGTGAAGCTTGCCCTCTGTCTACGGA 2633
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   1252 TGATGGCTGCAACTTCCACTTCTGTGGGAGAGCGCGGCTGCTTGCCCGCTCTGCTCAGT 1311

QY  2634 GCATGACTTCCATGAGATTGAGGGAGCCCTGCAAGAGAGGATTTTCAGGAAACCTTGTATGT 2693
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   1312 GGCTGACTACCATGCTATCGTCAAGCAGTGTGTGGTGGGATCCAGAAGACTACTTACGT 1371

QY  2694 GTGGAATGAACCTAAATGGTGCATTAAGGGAATTTCTTTCCTTGAGAAAAAGTTGGCAAC 2753
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   1372 GTGGCGAGAACCCCAAGCTATGCTCTGTGGCAATTTCTCTGCTGAGCAGAGATCAACCAT 1431

QY  2754 CTGTGAAACGGTTGACTTTTGGCTGAAGGTGGAGCGGCTGTGGAGCTTTTACTGCCCGT 2813
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   1432 CTGAAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCAACCTGTACTGCCAT 1491

QY  2814 TTTGCTGGTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATCAAAAACCTGGAATACAAATA 2873
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   1492 CCTGCTCACCGTCTTGACCTGCTACTTTTGGAAAAAGAAATCAAAAACCTAGAGTACAAGTA 1551
```



/gene="WUGSC:H\_RG331C24.1"  
/note="match to Q14832 (PID:g1171564); H\_RG331C24.1"  
/codon\_start=1  
/product="metabotropic glutamate receptor type 3 (mGluR3)"  
/protein\_id="AAC60379.1"  
/db\_xref="GI:2078454"  
translation="APFNPKNKADSIIVKFDTFDGMGRYNVFNQNVGGKYSYLKVGH  
WAETLSLDVNSIHSRNSVPTSQCDPCAPNEMKNMQPGDVCCWICIPCPEYEYLADE  
FTCMDCGGQWPTADLTGCDLPEDYIRWEDAWAIGPVTIACLGFMCTCMVVTVFIKH  
NNTPLVKASGRELCYILLEGVGLSYCMTFFFIAPKPSPVICALRRLGLGSSFAICYAL  
LTKNCIARIIFDGVKNGAQRPKFISPSQVFICIGLILVQIVMSVWLILEAPGTRRY  
TLAEKRETVILKCNVKDSSMLISLTYDVILVILCTVYAFKTRKCPENFNEAKFIGFTM  
YTTCTIIWLAFLPIFYVTSSDYRVQTTTMCISVLSLGFVVLGCLFAPKVHIILFQPQKN  
VVTHRLHLNRFVSVGTGTTYQSASSASTYVPTVCNGREVLDSTTSSL"  
22949..23079  
/rpt\_family="L2"  
23798..24042  
/rpt\_family="L1"  
24919..25091  
/rpt\_family="MERL\_type"  
25657..25937  
/gene="WUGSC:H\_RG331C24.1"  
/db\_xref="GI:1113596"  
26579..26649  
/rpt\_family="L2"  
26771..26919  
/rpt\_family="MERL\_type"  
27205..27582  
/rpt\_family="L2"  
27771..28069  
/rpt\_family="Alu"  
30158..30625  
/rpt\_family="L2"  
30909..30980  
/rpt\_family="L2"  
33325..33424  
/gene="WUGSC:H\_RG331C24.1"  
/note="match to EST AA670430 (NID:g2631929) ad20g04.sl"  
complement(33345..33424)  
/note="match to EST (NID:g1568674)"  
33557..33660  
/rpt\_family="L2"  
34927..35033  
/rpt\_family="MIR"  
35868..35981  
/rpt\_family="L2"  
36603..36746  
/rpt\_family="MaLR"  
37496..37634  
/rpt\_family="MERL\_type"  
39042..39351  
/rpt\_family="Alu"  
39357..39479  
/rpt\_family="L1"  
40234..40486  
/rpt\_family="L2"  
40835..41135  
/rpt\_family="Alu"  
41202..41565  
/rpt\_family="MaLR"  
41844..42141  
/rpt\_family="Alu"  
44395..45111  
/rpt\_family="L2"  
46236..46352  
/rpt\_family="MaLR"  
46911..47118  
/rpt\_family="L1"  
complement(47141..47518)

Query Match 14.6%; Score 518.4; DB 9; Length 136150;  
Best Local Similarity 99.8%; Pred. No. 2.4e-110;  
Matches 519; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3023 AGGAAAAAGAACCATTTTGAATCTGTTCAACTGAAAAACCTCAAGATCCCCAAATATAT 3082  
|||||  
Db 63412 AGGAAAAAGAACCATTTTGAATCTGTTCAACTGAAAAACCTCAAGATCCCCAAATATAT 63353  
QY 3083 GAAGACACAGTCTGTAGCCCTTGAGACTAATGAACAAAGAAAACCTGCTCTAGTTTACAG 3142  
|||||  
Db 63352 GAAGACACAGTCTGTAGCCCTTGAGACTAATGAACAAAGAAAACCTGCTCTAGTTTACAG 63293  
QY 3143 GACCATATTTTAGGGTCTGTCCCTCATACCTGTGCACATTGGTGATCTCACAGAGGAGGCC 3202  
|||||  
Db 63292 GACCATATTTTAGGGTCTGTCCCTCATACCTGTGCACATTGGTGATCTCACAGAGGAGGCC 63233  
QY 3203 ATGCCCGTGAAGAGGAGAGATTGAAACATTTGATTGCCCTTATCACATGGTCAAGTAC 3262  
|||||  
Db 63232 ATGCCCGTGAAGAGGAGAGATTGAAACATTTGATTGCCCTTATCACATGGTCAAGTAC 63173  
QY 3263 CTTGCCAAATAAAGGAAAGCAATGATTTGGGTCTCAACTGAAGCTCAAGCTCAACTCAG 3322  
|||||  
Db 63172 CTTGCCAAATAAAGGAAAGCAATGATTTGGGTCTCAACTGAAGCTCAAGCTCAACTCAG 63113  
QY 3323 GAAGAGATTTATCTGTATATACACATAACTGAAACCAAGTTTAAGCCCAATGCACCT 3382  
|||||  
Db 63112 GAAGAGATTTATCTGTATATACACATAACTGAAACCAAGTTTAAGCCCAATGCACCT 63053  
QY 3383 GCTGATGATGCCCATATAATTAATGGGTAACCTTTTATCTTTTATGATGTCTACATAACAA 3442  
|||||  
Db 63052 GCTGATGATGCCCATATAATTAATGGGTAACCTTTTATCTTTATGATGTCTACATAACAA 62993  
QY 3443 GTGTGATTTGGAAGGCACATGTGAGCATATGCATTATGATCCAATTTATTTTCTTT 3502  
|||||  
Db 62992 GTGTGATTTGGAAGGCACATGTGAGCATATGCATTATGATCCAATTTATTTTCTTT 62933  
QY 3503 GTTTATATTTTGGGGAAAAATTAATAATTTTAAAGGTAAA 3542  
|||||  
Db 62932 GTTTATATTTTGGGGAAAAATTAATAATTTTAAAGGTATA 62893

RESULT 15  
AX336856/c  
LOCUS AX336856 576 bp DNA linear PAT 09-JAN-2002  
DEFINITION Sequence 7365 from Patent WO0194629.  
ACCESSION AX336856  
VERSION AX336856.1 GI:18127575  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,  
Horrigan,S., Soppet,D.R. and Weaver,Z.  
TITLE Cancer gene determination and therapeutic screening using signature  
gene sets  
JOURNAL Patent: WO 0194629-A 7365 13-DEC-2001;  
Avalon Pharmaceuticals (US)  
FEATURES  
Source  
1..576  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 171 a 112 c 108 g 182 t 3 others  
ORIGIN

Query Match 14.2%; Score 504; DB 6; Length 576;  
Best Local Similarity 94.4%; Pred. No. 1.7e-107;  
Matches 543; Conservative 0; Mismatches 28; Indels 4; Gaps 2;  
QY 2970 TGTATATTCCTAATAAAGCAGTCACTACTAGGAAAACTCAAATCTTTGGCAACCAAGGAAAA 3029  
|||||.|||||  
Db 575 TGTATTCCTAATAAAGCAGTCACTACTAGGAAAACTCAAATCTTTGGCAACCAAGGAAAA 516  
QY 3030 AGAAGA---CCATTTTGAATCTGTT-CAACTGAAAAACCTCAAGATCCCCAAATATATGAA 3085  
|||||

```
Db 515 AAAGGAGACCCATTTTGAATCTGTTCCTCAACTGAAACCCCTCAAGATCCCCAAATATATGAA 456
QY 3086 GAGACAGTGTCTGAGCCTTGAGACTAATGAACAAGAAACCTGCTCTAGTTTACAGGAC 3145
Db 455 GAGACAGTCTGTAGCCCTTGAGACTAANTGANCAGGAAACCTGCTCTAGTTTACAGGAC 396
QY 3146 CATATTTTAGGGTCTGTCTCATACCTGTGCACATTTGGTGATCTCACAGAGGAGGCCCATG 3205
Db 395 CATATTTTAGGGTCTGTCTCATACCTGTGCACATTTGGTGATCTCACAGAGGAGGCCCATG 336
QY 3206 CCGCTGAAAAGGGAAGGAGATTGAAACATTTTGATTGCCCTTATCACATGGTCAAGTACCTT 3265
Db 335 CCCCTGAAAAGGGAAGGAGATTGAAACATTTTGATTGCCCTTATCACATGGTCAAGTACCTT 276
QY 3266 GCCAAATAAAGGAAAGCAAAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAA 3325
Db 275 GCCAAATAAAGGAAAGCAAAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAA 216
QY 3326 GAGATTTATCTGTATATACACATAACTGAAACCAAGTTTAAGCCCCACCAATGCACCTGCT 3385
Db 215 GAGATTTATCTGTATATACACATAACTGAAACCAAGTTTAAGCCCCACCAATGCACCTGCT 156
QY 3386 GATGCATGCCATATAATTAATGGGTAACTTTATCTTTATGATGTCTACATAACAAGTG 3445
Db 155 GATGCATGCCATATAATTAATGGGTAACTTTATCTTTATGATGTCTACATAACAAGTG 96
QY 3446 TGATTTGGAAGGCACATGTGAGCATATGCATTTATGATCCAAATTTATGTTTCTTTTGT 3505
Db 95 TGATTTGGAAGGCACATGTGAGCATATGCATTTATGATCCAAATTTATGTTTCTTTTGT 36
QY 3506 TATATTTTGGGGAAAAATTAATAATTTTAAAGGTA 3540
Db 35 TATATTTTGGGGAAAAATTAATAATTTTAAAGGTA 1
```



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:40:59 ; Search time 18.8729 Seconds  
(without alignments)  
4695.671 Million cell updates/sec

Title: US-10-073-333A-2  
Perfect score: 963  
Sequence: 1 MLFRARGPVRRGWRGPAEA.....TCYFWKKNQKKKKTILNLFN 963

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 349150 seqs, 92025710 residues

Word size : 0 349150

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	963	100.0	963	9	US-10-140-164-2	Sequence 2, Appli
2	953	99.0	1027	9	US-10-140-164-4	Sequence 4, Appli
3	271	28.1	464	9	US-10-002-050-20	Sequence 20, Appl
4	271	28.1	464	9	US-10-002-304-20	Sequence 20, Appl
5	271	28.1	464	12	US-10-003-152-20	Sequence 20, Appl
6	232	24.1	411	9	US-10-002-050-10	Sequence 10, Appl
7	232	24.1	411	9	US-10-002-304-10	Sequence 10, Appl
8	232	24.1	411	12	US-10-003-152-10	Sequence 10, Appl
9	81	8.4	81	10	US-09-864-761-39769	Sequence 39769, A
10	74	7.7	74	9	US-10-140-164-32	Sequence 32, Appl
11	74	7.7	74	9	US-10-140-164-61	Sequence 61, Appl
12	64	6.6	64	10	US-09-864-761-47095	Sequence 47095, A
13	60	6.2	60	10	US-09-864-761-39057	Sequence 39057, A
14	15	1.6	1013	9	US-10-028-072-38	Sequence 38, Appl
15	15	1.6	1013	9	US-10-121-049-38	Sequence 38, Appl
16	15	1.6	1013	9	US-10-123-904-38	Sequence 38, Appl
17	15	1.6	1013	9	US-10-140-470-38	Sequence 38, Appl
18	15	1.6	1013	9	US-10-175-746-38	Sequence 38, Appl
19	15	1.6	1013	9	US-10-176-918-38	Sequence 38, Appl

20	15	1.6	1013	9	US-10-176-921-38	Sequence 38, Appl
21	15	1.6	1013	9	US-10-137-865-38	Sequence 38, Appl
22	15	1.6	1013	9	US-10-140-474-38	Sequence 38, Appl
23	15	1.6	1013	9	US-10-142-431-38	Sequence 38, Appl
24	15	1.6	1013	9	US-10-143-114-38	Sequence 38, Appl
25	15	1.6	1013	9	US-10-140-002-38	Sequence 38, Appl
26	15	1.6	1013	9	US-10-142-419-38	Sequence 38, Appl
27	15	1.6	1013	9	US-10-123-262-38	Sequence 38, Appl
28	15	1.6	1013	9	US-10-142-423-38	Sequence 38, Appl
29	15	1.6	1013	9	US-10-121-050-38	Sequence 38, Appl
30	15	1.6	1013	9	US-10-141-755-38	Sequence 38, Appl
31	15	1.6	1013	9	US-10-143-032-38	Sequence 38, Appl
32	15	1.6	1013	9	US-10-123-108-38	Sequence 38, Appl
33	15	1.6	1013	9	US-10-123-236-38	Sequence 38, Appl
34	15	1.6	1013	9	US-10-123-261-38	Sequence 38, Appl
35	15	1.6	1013	9	US-10-140-921-38	Sequence 38, Appl
36	15	1.6	1013	9	US-10-140-928-38	Sequence 38, Appl
37	15	1.6	1013	9	US-10-121-045-38	Sequence 38, Appl
38	15	1.6	1013	9	US-10-123-292-38	Sequence 38, Appl
39	15	1.6	1013	9	US-10-123-903-38	Sequence 38, Appl
40	15	1.6	1013	9	US-10-124-819-38	Sequence 38, Appl
41	15	1.6	1013	9	US-10-124-822-38	Sequence 38, Appl
42	15	1.6	1013	9	US-10-140-925-38	Sequence 38, Appl
43	15	1.6	1013	9	US-10-160-498-38	Sequence 38, Appl
44	15	1.6	1013	9	US-10-121-041-38	Sequence 38, Appl
45	15	1.6	1013	9	US-10-121-043-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1  
US-10-140-164-2  
; Sequence 2, Application US/10140164  
; Publication No. US20030072736A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker et al.  
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16  
; FILE REFERENCE: PF514C1  
; CURRENT APPLICATION NUMBER: US/10/140,164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 963  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-140-164-2

Query Match 100.0%; Score 963; DB 9; Length 963;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLFRARGPVRRGWRGPAEAPRRGRSPWPSPAWICCWALAGCQAAWAGDLPSSSRPLPP 60  
|||||  
Db 1 MLFRARGPVRRGWRGPAEAPRRGRSPWPSPAWICCWALAGCQAAWAGDLPSSSRPLPP 60

QY 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120  
Db 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120  
QY 121 SKCGEGTYSLGSGIKFDEWDELPAFNSNIATFMDTVVGPSPDSRPDGCNNSSWIPRGNYIE 180  
Db 121 SKCGEGTYSLGSGIKFDEWDELPAFNSNIATFMDTVVGPSPDSRPDGCNNSSWIPRGNYIE 180  
QY 181 SNRDDCTVSLIYAVHLKSGYVFFEQYVDNNIFFEQNDQCQEMDTTDDKWVKLTDN 240  
Db 181 SNRDDCTVSLIYAVHLKSGYVFFEQYVDNNIFFEQNDQCQEMDTTDDKWVKLTDN 240  
QY 241 GEWGSVLMKSGTNILYWRRTGILMGSKAVKPVLVKNITIEGVAYTSECFCPCPGTFSN 300  
Db 241 GEWGSVLMKSGTNILYWRRTGILMGSKAVKPVLVKNITIEGVAYTSECFCPCPGTFSN 300  
QY 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360  
Db 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360  
QY 361 KTOIMYKWIIEPKICREDLTDAIRLPPSGEKKDCPPCNPFGYNNGSSSCHPCPPGTFSDGT 420  
Db 361 KTOIMYKWIIEPKICREDLTDAIRLPPSGEKKDCPPCNPFGYNNGSSSCHPCPPGTFSDGT 420  
QY 421 KECRPCPAGTEPALGFYKWNVLPGNMKTSFCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480  
Db 421 KECRPCPAGTEPALGFYKWNVLPGNMKTSFCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480  
QY 481 DNDYLILNLHIPGFKPPTSMGTATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540  
Db 481 DNDYLILNLHIPGFKPPTSMGTATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540  
QY 541 SWGGTKEKQAYTHIIFKNATFTTWAQFQRTNQODNRRFINDMVKIYSITATNAVGVAS 600  
Db 541 SWGGTKEKQAYTHIIFKNATFTTWAQFQRTNQODNRRFINDMVKIYSITATNAVGVAS 600  
QY 601 SCRACALGSEQSGSSCVPCPPGHYIEKETNQCKECPDPTLSIHQVYGKEACIPCGPGSK 660  
Db 601 SCRACALGSEQSGSSCVPCPPGHYIEKETNQCKECPDPTLSIHQVYGKEACIPCGPGSK 660  
QY 661 NNQDHSVCSYDCFFYHEKENQILHYDFSNLSSVGLMNGPSTFKGTKYFHHFNISLCGH 720  
Db 661 NNQDHSVCSYDCFFYHEKENQILHYDFSNLSSVGLMNGPSTFKGTKYFHHFNISLCGH 720  
QY 721 EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFCVQSTIIPSESKGFRAALSSQSIILA 780  
Db 721 EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFCVQSTIIPSESKGFRAALSSQSIILA 780  
QY 781 DTFIGVTETLKNINIKEDMFPVPTSOIPDVHFFYKSTATTSCINGRSTAVKMRNPT 840  
Db 781 DTFIGVTETLKNINIKEDMFPVPTSOIPDVHFFYKSTATTSCINGRSTAVKMRNPT 840  
QY 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAECPLCTEHDFHEIGACKRGFOETLYVWN 900  
Db 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAECPLCTEHDFHEIGACKRGFOETLYVWN 900  
QY 901 EPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCTCYFWKKNQKKKTILN 960  
Db 901 EPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCTCYFWKKNQKKKTILN 960  
QY 961 LFN 963  
Db 961 LFN 963

RESULT 2

US-10-140-164-4  
; Sequence 4, Application US/10140164  
; Publication No. US2003007236A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker et al.  
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16  
; FILE REFERENCE: PF514C1

; CURRENT APPLICATION NUMBER: US/10/140,164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 1027  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-140-164-4

Query Match 99.0%; Score 953; DB 9; Length 1027;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 953; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFRARGPVRGRGWGRPAEAPRRGRSPWPSPAWICCWALAGCOAAWAGDLPSSSSRPLPP 60  
Db 1 MLFRARGPVRGRGWGRPAEAPRRGRSPWPSPAWICCWALAGCOAAWAGDLPSSSSRPLPP 60  
QY 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120  
Db 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120  
QY 121 SKCGEGTYSLGSGIKFDEWDELPAFNSNIATFMDTVVGPSPDSRPDGCNNSSWIPRGNYIE 180  
Db 121 SKCGEGTYSLGSGIKFDEWDELPAFNSNIATFMDTVVGPSPDSRPDGCNNSSWIPRGNYIE 180  
QY 181 SNRDDCTVSLIYAVHLKSGYVFFEQYVDNNIFFEQNDQCQEMDTTDDKWVKLTDN 240  
Db 181 SNRDDCTVSLIYAVHLKSGYVFFEQYVDNNIFFEQNDQCQEMDTTDDKWVKLTDN 240  
QY 241 GEWGSVLMKSGTNILYWRRTGILMGSKAVKPVLVKNITIEGVAYTSECFCPCPGTFSN 300  
Db 241 GEWGSVLMKSGTNILYWRRTGILMGSKAVKPVLVKNITIEGVAYTSECFCPCPGTFSN 300  
QY 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360  
Db 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360  
QY 361 KTOIMYKWIIEPKICREDLTDAIRLPPSGEKKDCPPCNPFGYNNGSSSCHPCPPGTFSDGT 420  
Db 361 KTOIMYKWIIEPKICREDLTDAIRLPPSGEKKDCPPCNPFGYNNGSSSCHPCPPGTFSDGT 420  
QY 421 KECRPCPAGTEPALGFYKWNVLPGNMKTSFCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480  
Db 421 KECRPCPAGTEPALGFYKWNVLPGNMKTSFCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480  
QY 481 DNDYLILNLHIPGFKPPTSMGTATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540  
Db 481 DNDYLILNLHIPGFKPPTSMGTATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540  
QY 541 SWGGTKEKQAYTHIIFKNATFTTWAQFQRTNQODNRRFINDMVKIYSITATNAVGVAS 600  
Db 541 SWGGTKEKQAYTHIIFKNATFTTWAQFQRTNQODNRRFINDMVKIYSITATNAVGVAS 600  
QY 601 SCRACALGSEQSGSSCVPCPPGHYIEKETNQCKECPDPTLSIHQVYGKEACIPCGPGSK 660  
Db 601 SCRACALGSEQSGSSCVPCPPGHYIEKETNQCKECPDPTLSIHQVYGKEACIPCGPGSK 660

Qy 661 NNQDHSVYSDCFFYHEKENQILHYDFSNLSSVGLMNGPSFTSKGTYKYPHFHNISLCGH 720  
Db 661 NNQDHSVYSDCFFYHEKENQILHYDFSNLSSVGLMNGPSFTSKGTYKYPHFHNISLCGH 720  
Qy 721 EGKMACTNNITDFTVKEIVAGSDDYTNLVGAFVCQSTIIPSESKGFRALSSQSIILA 780  
Db 721 EGKMACTNNITDFTVKEIVAGSDDYTNLVGAFVCQSTIIPSESKGFRALSSQSIILA 780  
Qy 781 DTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSATTATTCINGRSTAVKMRNPT 840  
Db 781 DTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSATTATTCINGRSTAVKMRNPT 840  
Qy 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAECPLCTEHDHFHEIEGACKRGFQETLYVWN 900  
Db 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAECPLCTEHDHFHEIEGACKRGFQETLYVWN 900  
Qy 901 EPKWCIGISLPEKKLATCETVDFWLKVGAGVGAFATVLLVALTCYFWKKNQK 953  
Db 901 EPKWCIGISLPEKKLATCETVDFWLKVGAGVGAFATVLLVALTCYFWKKNQK 953

RESULT 3  
US-10-002-050-20  
; Sequence 20, Application US/10002050  
; Publication No. US20030032095A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: No. US20030032095A1 Nucleic Acid Sequences Encoding Human Semaphorin 3A  
; FILE REFERENCE: 15966-554 Cura-54 CON-S14  
; CURRENT APPLICATION NUMBER: US/10/002,050  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 464  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-002-050-20

Query Match 28.1%; Score 271; DB 9; Length 464;  
Best Local Similarity 99.6%; Pred. No. 1.6e-257;  
Matches 451; Conservative 0; Mismatches 0; Indels 2; Gaps 1;  
Qy 115 MKNQVCSKCGEGTYSLSGSIKFDWEDELPAFGFSNIATFMDTVVGPDSRDPDGCNNSSWIP 174  
Db 1 MKNQVCSKCGEGTYSLSGSIKFDWEDELPAFGFSNIATFMDTVVGPDSRDPDGCNNSSWIP 60  
Qy 175 RGNYESNRDDCTVSLIYAVHLKKSQYVFFYQYVDNNIFFEFFFIONDQCQEMDTTTDKW 234  
Db 61 RGNYESNRDDCTVSLIYAVHLKKSQYVFFYQYVDNNIFFEFFFIONDQCQEMDTTTDKW 120  
Qy 235 VKLTDNGEWGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFFCK 294  
Db 121 VKLTDNGEWGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFFCK 180  
Qy 295 PGTFSENKPGSFNCQVCPRNTYSEKGAKECIRCKDDSOFS--GSSECTERPPCTTKDYFQI 352  
Db 181 PGTFSENKPGSFNCQVCPRNTYSEKGAKECIRCKDDSOFS--GSSECTERPPCTTKDYFQI 240  
Qy 353 HTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPNGFYNNNGSSSCHPCP 412  
Db 181 PGTFSENKPGSFNCQVCPRNTYSEKGAKECIRCKDDSOFS--GSSECTERPPCTTKDYFQI 240  
Qy 353 HTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPNGFYNNNGSSSCHPCP 412  
Db 241 HTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPNGFYNNNGSSSCHPCP 300

Qy 413 PGTFSDGTEKCRPCPAGTEPALGFEPYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 472  
Db 301 PGTFSDGTEKCRPCPAGTEPALGFEPYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 360  
Qy 473 IQSGAGGSDNDYLILNLHIPGFKPPTSMGTATGSELGRITTFVFETLCSADCVLVFMVDIN 532  
Db 361 IQSGAGGSDNDYLILNLHIPGFKPPTSMGTATGSELGRITTFVFETLCSADCVLVFMVDIN 420  
Qy 533 RKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 565  
Db 421 RKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 453  
RESULT 4  
US-10-002-304-20  
; Sequence 20, Application US/10002304  
; Publication No. US20030036185A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby  
; FILE REFERENCE: 15966-554 Cura-54 CON-S8  
; CURRENT APPLICATION NUMBER: US/10/002,304  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 20  
; LENGTH: 464  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-002-304-20

Query Match 28.1%; Score 271; DB 9; Length 464;  
Best Local Similarity 99.6%; Pred. No. 1.6e-257;  
Matches 451; Conservative 0; Mismatches 0; Indels 2; Gaps 1;  
Qy 115 MKNQVCSKCGEGTYSLSGSIKFDWEDELPAFGFSNIATFMDTVVGPDSRDPDGCNNSSWIP 174  
Db 1 MKNQVCSKCGEGTYSLSGSIKFDWEDELPAFGFSNIATFMDTVVGPDSRDPDGCNNSSWIP 60  
Qy 175 RGNYESNRDDCTVSLIYAVHLKKSQYVFFYQYVDNNIFFEFFFIONDQCQEMDTTTDKW 234  
Db 61 RGNYESNRDDCTVSLIYAVHLKKSQYVFFYQYVDNNIFFEFFFIONDQCQEMDTTTDKW 120  
Qy 235 VKLTDNGEWGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFFCK 294  
Db 121 VKLTDNGEWGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFFCK 180  
Qy 295 PGTFSENKPGSFNCQVCPRNTYSEKGAKECIRCKDDSOFS--GSSECTERPPCTTKDYFQI 352  
Db 181 PGTFSENKPGSFNCQVCPRNTYSEKGAKECIRCKDDSOFS--GSSECTERPPCTTKDYFQI 240  
Qy 353 HTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPNGFYNNNGSSSCHPCP 412  
Db 241 HTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPNGFYNNNGSSSCHPCP 300  
Qy 413 PGTFSDGTEKCRPCPAGTEPALGFEPYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 472  
Db 301 PGTFSDGTEKCRPCPAGTEPALGFEPYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 360  
Qy 473 IQSGAGGSDNDYLILNLHIPGFKPPTSMGTATGSELGRITTFVFETLCSADCVLVFMVDIN 532  
Db 361 IQSGAGGSDNDYLILNLHIPGFKPPTSMGTATGSELGRITTFVFETLCSADCVLVFMVDIN 420  
Qy 533 RKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 565



Db	421	RKSTNVVSWGTTKEKQAYTHIFKNATFTTW	453
RESULT 5			
US-10-003-152-20			
; Sequence 20, Application US/10003152			
; Patent No. US20020151494A1			
; GENERAL INFORMATION:			
; APPLICANT: Shimkets, Richard			
; APPLICANT: Fernandes, Elma			
; APPLICANT: Vernet, Corine			
; APPLICANT: Yang, Meijia			
; APPLICANT: Boldog, Ferenc			
; APPLICANT: Herrmann, John			
; TITLE OF INVENTION: No. US20020151494A1el Amino Acid Sequences for Human Semaphorin-1			
; FILE REFERENCE: 15966-554 Cura-54 CON-S12			
; CURRENT APPLICATION NUMBER: US/10/003,152			
; CURRENT FILING DATE: 2001-11-02			
; PRIOR APPLICATION NUMBER: 09/604,286			
; PRIOR FILING DATE: 2000-06-22			
; PRIOR APPLICATION NUMBER: 60/140,584			
; PRIOR FILING DATE: 1999-06-23			
; NUMBER OF SEQ ID NOS: 49			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 20			
; LENGTH: 464			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-10-003-152-20			
Query Match 28.1%; Score 271; DB 12; Length 464;			
Best Local Similarity 99.6%; Pred. No. 1.6e-257;			
Matches 451; Conservative 0; Mismatches 0; Indels 2; Gaps 1;			
QY	115	MKNQVCSKCGEGTSLIYAVHLKKGSGIKFDEWDELPAFNSNIATFMDTVVGPDSRDPDGCNNSSWIP	174
Db	1	MKNQVCSKCGEGTSLIYAVHLKKGSGIKFDEWDELPAFNSNIATFMDTVVGPDSRDPDGCNNSSWIP	60
QY	175	RGNIESNRDDCTVSLIYAVHLKKGSGYVFEFYQYVDNNIFFEFIQNDQCQEMDTTDDKW	234
Db	61	RGNIESNRDDCTVSLIYAVHLKKGSGYVFEFYQYVDNNIFFEFIQNDQCQEMDTTDDKW	120
QY	235	VKLTDNWEGSHSVMLKSGTNILYWRRTGILMGSKAVKPVLVKNITIEGVAYTSECFPCK	294
Db	121	VKLTDNWEGSHSVMLKSGTNILYWRRTGILMGSKAVKPVLVKNITIEGVAYTSECFPCK	180
QY	295	PGTFSNKPQSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTERPPCTTKDYFQI	352
Db	181	PGTFSNKPQSFNCQVCPRNTYSEKGAKECIRCKDDSQFSSEGSSECTERPPCTTKDYFQI	240
QY	353	HTPCDEEGTKQIMYKWIIEPKICREDLTDAIRLPPSGEKKDCPPCNPFGYNNNGSSSCHPCP	412
Db	241	HTPCDEEGTKQIMYKWIIEPKICREDLTDAIRLPPSGEKKDCPPCNPFGYNNNGSSSCHPCP	300
QY	413	PGTFSDGTECRPCPAGTEPALGFYKWWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH	472
Db	301	PGTFSDGTECRPCPAGTEPALGFYKWWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH	360
QY	473	IQSGAGGSDNDYLILNLHIPGFKPPTSMGTATGSELGRITFVFETLCSADCVLYFMVDIN	532
Db	361	IQSGAGGSDNDYLILNLHIPGFKPPTSMGTATGSELGRITFVFETLCSADCVLYFMVDIN	420
QY	533	RKSTNVVSWGTTKEKQAYTHIFKNATFTTW	565
Db	421	RKSTNVVSWGTTKEKQAYTHIFKNATFTTW	453
RESULT 6			
US-10-002-050-10			
; Sequence 10, Application US/10002050			
; Publication No. US20030032095A1			
; GENERAL INFORMATION:			
; APPLICANT: Shimkets, Richard			
; APPLICANT: Fernandes, Elma			
; APPLICANT: Vernet, Corine			
; APPLICANT: Yang, Meijia			
; APPLICANT: Boldog, Ferenc			
; APPLICANT: Herrmann, John			
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby			
; FILE REFERENCE: 15966-554 Cura-54 CON-S8			
; CURRENT APPLICATION NUMBER: US/10/002,304			
; CURRENT FILING DATE: 2001-11-02			
; PRIOR APPLICATION NUMBER: 09/604,286			
; PRIOR FILING DATE: 2000-06-22			
; PRIOR APPLICATION NUMBER: 60/140,584			
; PRIOR FILING DATE: 1999-06-23			
; NUMBER OF SEQ ID NOS: 49			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 10			
; LENGTH: 411			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-10-002-304-10			
Query Match 24.1%; Score 232; DB 9; Length 411;			
Best Local Similarity 100.0%; Pred. No. 2.9e-219;			
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			



QY 334 GSSECTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIIEPKICREDLTDAIRLPPSGEKKDC 393  
Db 169 GSSECTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIIEPKICREDLTDAIRLPPSGEKKDC 228  
QY 394 PPCNPGFYNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCF 453  
Db 229 PPCNPGFYNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCF 288  
QY 454 NVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMGTGATGSELGRITF 513  
Db 289 NVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMGTGATGSELGRITF 348  
QY 514 VFETLCSADCVLVFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 565  
Db 349 VFETLCSADCVLVFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 400  
RESULT 8  
US-10-003-152-10  
; Sequence 10, Application US/10003152  
; Patent No. US20020151494A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: No. US20020151494A1 Amino Acid Sequences for Human Semaphorin-I  
; FILE REFERENCE: 15966-554 Cura-54 CON-S12  
; CURRENT APPLICATION NUMBER: US/10/003,152  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 411  
; TYPE: PRT  
; ORGANISM: Homo sapiens.  
US-10-003-152-10  
Query Match 24.1%; Score 232; DB 12; Length 411;  
Best Local Similarity 100.0%; Pred. No. 2.9e-219;  
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 334 GSSECTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIIEPKICREDLTDAIRLPPSGEKKDC 393  
Db 169 GSSECTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIIEPKICREDLTDAIRLPPSGEKKDC 228  
QY 394 PPCNPGFYNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCF 453  
Db 229 PPCNPGFYNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCF 288  
QY 454 NVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMGTGATGSELGRITF 513  
Db 289 NVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMGTGATGSELGRITF 348  
QY 514 VFETLCSADCVLVFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 565  
Db 349 VFETLCSADCVLVFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 400  
RESULT 9  
US-09-864-761-39769  
; Sequence 39769, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aeomica-x-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 39769  
; LENGTH: 81  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC002081.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2  
; OTHER INFORMATION: EST\_HUMAN HIT: AW954806.1, EVALUE 5.00e-45  
; OTHER INFORMATION: SWISSPROT HIT: P01267, EVALUE 3.00e-03  
US-09-864-761-39769  
Query Match 8.4%; Score 81; DB 10; Length 81;  
Best Local Similarity 100.0%; Pred. No. 1.5e-71;  
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 578 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSVCPPPGHYIEKETNQCKECP 637  
Db 1 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSVCPPPGHYIEKETNQCKECP 60  
QY 638 DTYLSIHQVYGKEACIPCGPG 658  
Db 61 DTYLSIHQVYGKEACIPCGPG 81

RESULT 10  
US-10-140-164-32  
; Sequence 32, Application US/10140164  
; Publication No. US20030072736A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker et al.  
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16  
; FILE REFERENCE: PF514C1  
; CURRENT APPLICATION NUMBER: US/10/140,164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 32  
; LENGTH: 74  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-140-164-32

Query Match 7.7%; Score 74; DB 9; Length 74;  
Best Local Similarity 100.0%; Pred. No. 1e-64;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 CTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIETKICREDLTDAIRLPPSGEKKDCPPCN 397  
|||||  
Db 1 CTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIETKICREDLTDAIRLPPSGEKKDCPPCN 60  
QY 398 PGFYNNGSSSCHPC 411  
|||||  
Db 61 PGFYNNGSSSCHPC 74

RESULT 11  
US-10-140-164-61  
; Sequence 61, Application US/10140164  
; Publication No. US20030072736A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker et al.  
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16  
; FILE REFERENCE: PF514C1  
; CURRENT APPLICATION NUMBER: US/10/140,164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18

; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 61  
; LENGTH: 74  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-140-164-61.

Query Match 7.7%; Score 74; DB 9; Length 74;  
Best Local Similarity 100.0%; Pred. No. 1e-64;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 CTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIETKICREDLTDAIRLPPSGEKKDCPPCN 397  
|||||  
Db 1 CTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIETKICREDLTDAIRLPPSGEKKDCPPCN 60  
QY 398 PGFYNNGSSSCHPC 411  
|||||  
Db 61 PGFYNNGSSSCHPC 74

RESULT 12  
US-09-864-761-47095  
; Sequence 47095, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47095
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002081.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.55
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.52
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.49
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.49
; OTHER INFORMATION: SWISSPROT HIT: Q00019, EVALUE 8.90e-01
US-09-864-761-47095

Query Match 6.6%; Score 64; DB 10; Length 64;
Best Local Similarity 100.0%; Pred. No. 6.1e-55;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 FSCASGEYLEMKNQVCCKGEGTYSLGSGIKFDEWDELPAGFSNIATFMDTVVGPDSRP 164
|||||
Db 1 FSCASGEYLEMKNQVCCKGEGTYSLGSGIKFDEWDELPAGFSNIATFMDTVVGPDSRP 60

QY 165 DGCN 168
||||
Db 61 DGCN 64

RESULT 13
US-09-864-761-39057
; Sequence 39057, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39057
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002081.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
; OTHER INFORMATION: EST\_HUMAN HIT: AW966212.1, EVALUE 2.00e-10
; OTHER INFORMATION: SWISSPROT HIT: P21849, EVALUE 5.00e-03
US-09-864-761-39057

Query Match 6.2%; Score 60; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 4.9e-51;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 TQIMYKWIEPKICREDLTDAIRLPPSGEKKDCPPCNGSGSCHPCPPGTFSDGTK 421
|||||
Db 1 TQIMYKWIEPKICREDLTDAIRLPPSGEKKDCPPCNGSGSCHPCPPGTFSDGTK 60

RESULT 14
US-10-028-072-38
; Sequence 38, Application US/10028072
; Publication No. US20030004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang
; TITLE OF INVENTION:
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/028,072
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263









GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:23:58 ; Search time 13.0658 Seconds  
(without alignments)  
3056.959 Million cell updates/sec

Title: US-10-073-333A-2  
Perfect score: 963  
Sequence: 1 MLFRARGPVRGRGWGRPAEA.....TCYFWKKKQKKKTIILNFN 963

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	8	0.8	100	1	DEF6_HUMAN	Q01524 homo sapien
2	8	0.8	271	1	TNR4_RAT	P15725 rattus norv.
3	8	0.8	272	1	TNR4_MOUSE	P47741 mus musculu
4	8	0.8	293	1	VBLL_TGMV	P03566 tomato gold
5	8	0.8	300	1	TR6B_HUMAN	O95407 homo sapien
6	8	0.8	405	1	DCP2_PEA	P51851 pisum sativ
7	8	0.8	418	1	DCP1_TOBAC	P51845 nicotiana t
8	8	0.8	430	1	TRLT_HUMAN	Q96924 homo sapien
9	8	0.8	430	1	TRLT_MACFA	Q9n092 macaca fasc
10	8	0.8	585	1	DCP3_ORYSA	P51849 oryza sativ
11	8	0.8	603	1	DCP2_ORYSA	P51848 oryza sativ
12	8	0.8	671	1	KRP2_RAT	Q62909 rattus norv
13	8	0.8	718	1	MCAK_CRIGR	P70096 cricetulus
14	7	0.7	100	1	TAT_SIVCZ	P17285 chimpanzee
15	7	0.7	101	1	RK24_GUITH	O46905 guillardia
16	7	0.7	134	1	CRB2_BACHD	Q9k819 bacillus ha
17	7	0.7	141	1	NEUV_RANES	P11858 rana escule
18	7	0.7	145	1	MMS2_MVCTU	Q11170 mycobacteri
19	7	0.7	159	1	NEUV_BUFJA	P08163 bufo japoni
20	7	0.7	179	1	YCBQ_ECOLI	P75855 escherichia
21	7	0.7	179	1	YD85_METJA	Q58780 methanococc
22	7	0.7	200	1	RS8A_SCHPO	O14049 schizosacch
23	7	0.7	200	1	RS8B_SCHPO	Q9p7b2 schizosacch
24	7	0.7	203	1	Y296_METJA	Q57744 methanococc
25	7	0.7	206	1	Y93Q_HAEIN	P44077 haemophilus
26	7	0.7	208	1	RASM_MOUSE	O08989 mus musculu
27	7	0.7	208	1	RASM_RAT	P97538 rattus norv
28	7	0.7	208	1	YML6_YEAST	P40206 saccharomyc
29	7	0.7	208	1	YQED_BACSU	P54449 bacillus su
30	7	0.7	259	1	SPO7_YEAST	P18410 saccharomyc
31	7	0.7	270	1	PYRF_CANAL	P13649 candida alb
32	7	0.7	270	1	PYRF_CANDU	Q9c150 candida dub
33	7	0.7	277	1	ATND_XENLA	P21188 xenopus lae

34	7	0.7	283	1	TR14_HUMAN	Q92956 homo sapien
35	7	0.7	320	1	DNC_HUMAN	Q9hc21 homo sapien
36	7	0.7	335	1	FLIG_THEMA	Q9wy63 thermotoga
37	7	0.7	362	1	YD33_MYCLE	P53425 mycobacteri
38	7	0.7	365	1	WN14_HUMAN	O14904 homo sapien
39	7	0.7	418	1	YE87_SCHPO	O14303 schizosacch
40	7	0.7	420	1	YCEL_CAEEL	Q94175 caenorhabdi
41	7	0.7	437	1	YQEZ_BACSU	P54465 bacillus su
42	7	0.7	459	1	G33_RAT	P05432 rattus norv
43	7	0.7	463	1	YAD4_YEAST	P28003 saccharomyc
44	7	0.7	529	1	YPC1_CAEEL	Q11178 caenorhabdi
45	7	0.7	548	1	TRM1_SCHPO	Q9p804 schizosacch

ALIGNMENTS

RESULT 1						
DEF6_HUMAN						
ID	DEF6_HUMAN	STANDARD;	PRT;	100	AA.	
AC	Q01524;					
DT	01-JUL-1993 (Rel. 26, Created)					
DT	01-JUL-1993 (Rel. 26, Last sequence update)					
DT	15-JUN-2002 (Rel. 41, Last annotation update)					
DE	Defensin 6 precursor (Defensin, alpha 6).					
GN	DEFA6 OR DEF6.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=Intestine;					
RX	MEDLINE=93114459; PubMed=8417977;					
RA	Jones D.E., Bevins C.L.;					
RT	"Defensin-6 mRNA in human Paneth cells: implications for antimicrobial peptides in host defense of the human bowel.";					
RL	FEBS Lett. 315:187-192(1993).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=96223969; PubMed=8626737;					
RA	Mallow E.B., Harris A., Salzman N., Russell J.P.,					
RA	Deberardinis R.J., Ruchelli E., Bevins C.L.;					
RT	"Human enteric defensins. Gene structure and developmental expression.";					
RL	J. Biol. Chem. 271:4038-4045(1996).					
CC	-!- FUNCTION: HAS ANTIMICROBIAL ACTIVITY.					
CC	-!- TISSUE SPECIFICITY: PANETH CELLS OF THE SMALL INTESTINE.					
CC	-!- SIMILARITY: BELONGS TO THE CORTICOSTATIN/DEFENSIN FAMILY.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).					
CC	-----					
DR	EMBL; M98331; AAB59357.1;					
DR	EMBL; U33317; AAC50382.1; ALT_SEQ.					
DR	PIR; S27016; S27016.					
DR	Genew; HGNC:2765; DEFA6.					
DR	MIM; 600471;					
DR	InterPro; IPR002366; Defensin_alpha.					
DR	InterPro; IPR001271; Defensin_mammal.					
DR	Pfam; PF00323; defensins; 1.					
DR	Pfam; PF00879; Defensin_propep; 1.					
DR	SMART; SM00048; DEFSN; 1.					
DR	PROSITE; PS00269; DEFENSIN; 1.					
DR	Defensin; Antibiotic; Fungicide; Signal.					
KW	SIGNAL 1 19					
FT	PROPEP 20 ?65					
FT	CHAIN ?66 100					
FT	DEFENSIN 6.					

```
FT  DISULFID      72      99      BY SIMILARITY.
FT  DISULFID      74      88      BY SIMILARITY.
FT  DISULFID      78      98      BY SIMILARITY.
SQ  SEQUENCE    100 AA; 10975 MW;  EDF77E033DDCE2D5 CRC64;

Query Match          0.8%; Score 8; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  936 TAVLLVAL 943
    |||||
Db  8 TAVLLVAL 15

RESULT 2
TNR4_RAT
ID  TNR4_RAT      STANDARD;      PRT;      271 AA.
AC  P15725;
DT  01-APR-1990 (Rel. 14, Created)
DT  01-APR-1990 (Rel. 14, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Tumor necrosis factor receptor superfamily member 4 precursor (OX40L
DE  receptor) (OX40 antigen) (MRC OX40).
GN  TNFRSF4 OR TXGP1L OR OX40.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=T-cell;
RX  MEDLINE=90214614; PubMed=2157591;
RA  Mallett S., Fossum S., Barclay A.N.;
RT  "Characterization of the MRC OX40 antigen of activated CD4 positive T
RL  lymphocytes -- a molecule related to nerve growth factor receptor.";
RL  EMBO J. 9:1063-1068(1990).
CC  -!- FUNCTION: Receptor for TNFSF4/OX40L/GP34.
CC  -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC  -!- TISSUE SPECIFICITY: ACTIVATED T-CELLS.
CC  -!- SIMILARITY: CONTAINS 4 TNFR-CYS REPEATS.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL; X17037; CAA34897.1; -.
DR  PIR; S08036; S08036.
DR  PIR; S12783; S12783.
DR  HSSP; O14763; 1D4V.
DR  InterPro; IPR001368; TNFR_c6.
DR  Pfam; PF00020; TNFR_c6; 3.
DR  ProDom; PD000771; TNFR_c6; 1.
DR  SMART; SM00208; TNFR; 3.
DR  PROSITE; PS00652; TNFR_NGFR_1; 2.
DR  PROSITE; PS50050; TNFR_NGFR_2; 2.
KW  Receptor; Antigen; Transmembrane; Glycoprotein; Repeat; Signal.
FT  SIGNAL      1 19 POTENTIAL.
FT  CHAIN      20 271
FT  DOMAIN      20 210
FT  TRANSMEM    211 235
FT  DOMAIN      236 271
FT  REPEAT      25 60
FT  REPEAT      61 102
FT  REPEAT      103 123
FT  REPEAT      124 164
FT  REPEAT      26 37
FT  DISULFID     38 51
FT  DISULFID     41 59
FT  DISULFID     41 59
BY SIMILARITY.
```

```
FT  DISULFID      62      76      BY SIMILARITY.
FT  DISULFID      79      94      BY SIMILARITY.
FT  DISULFID      82     102      BY SIMILARITY.
FT  DISULFID     104     122      BY SIMILARITY.
FT  DISULFID     125     138      BY SIMILARITY.
FT  DISULFID     144     163      BY SIMILARITY.
FT  CARBOHYD     143     143      N-LINKED (GLCNAC....) (POTENTIAL).
SQ  SEQUENCE    271 AA; 29895 MW;  C06465136B16E821 CRC64;

Query Match          0.8%; Score 8; DB 1; Length 271;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  616 CVPCPPGH 623
    |||||
Db  122 CVPCPPGH 129

RESULT 3
TNR4_MOUSE
ID  TNR4_MOUSE      STANDARD;      PRT;      272 AA.
AC  P47741;
DT  01-FEB-1996 (Rel. 33, Created)
DT  01-FEB-1996 (Rel. 33, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Tumor necrosis factor receptor superfamily member 4 precursor (OX40L
DE  receptor) (OX40 antigen).
GN  TNFRSF4 OR TXGP1 OR OX40.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BALB/c;
RX  MEDLINE=94044750; PubMed=8228223;
RA  Calderhead D.M., Buhlmann J.E., van den Eertwegh A.J.,
RA  Claassen E., Noelle R.J., Fell H.;
RT  "Cloning of mouse OX40: a T cell activation marker that may mediate
RT  T-B cell interactions.";
RL  J. Immunol. 151:5261-5271(1993).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=95255413; PubMed=7737295;
RA  Birkeland M.L., Copeland N.G., Gilbert D.J., Jenkins N.A.,
RA  Barclay A.N.;
RT  "Gene structure and chromosomal localization of the mouse homologue
RT  of rat OX40 protein.";
RL  Eur. J. Immunol. 25:926-930(1995).
CC  -!- FUNCTION: Receptor for TNFSF4/OX40L/GP34.
CC  -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC  -!- SIMILARITY: CONTAINS 4 TNFR-CYS REPEATS.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL; Z21674; CAA79772.1; -.
DR  EMBL; X85214; CAA59476.1; -.
DR  HSSP; O14763; 1DOG.
DR  MGD; MGI:104512; Tnfrsf4.
DR  InterPro; IPR001368; TNFR_c6.
DR  Pfam; PF00020; TNFR_c6; 3.
DR  ProDom; PD000771; TNFR_c6; 1.
DR  SMART; SM00208; TNFR; 3.
DR  PROSITE; PS00652; TNFR_NGFR_1; 2.
DR  PROSITE; PS50050; TNFR_NGFR_2; 2.
KW  Receptor; Antigen; Transmembrane; Glycoprotein; Repeat; Signal.
FT  SIGNAL      1 19 POTENTIAL.
FT  CHAIN      20 271
```



FT CHAIN 20 272 TUMOR NECROSIS FACTOR RECEPTOR  
FT SUPERFAMILY MEMBER 4.  
FT DOMAIN 20 211 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 212 236 POTENTIAL.  
FT DOMAIN 237 272 CYTOPLASMIC (POTENTIAL).  
FT REPEAT 26 61 TNFR-CYS 1.  
FT REPEAT 62 103 TNFR-CYS 2.  
FT REPEAT 104 124 TNFR-CYS 3 (INCOMPLETE).  
FT REPEAT 125 165 TNFR-CYS 4.  
FT DISULFID 27 38 BY SIMILARITY.  
FT DISULFID 39 52 BY SIMILARITY.  
FT DISULFID 42 60 BY SIMILARITY.  
FT DISULFID 63 77 BY SIMILARITY.  
FT DISULFID 80 95 BY SIMILARITY.  
FT DISULFID 83 103 BY SIMILARITY.  
FT DISULFID 105 123 BY SIMILARITY.  
FT DISULFID 126 139 BY SIMILARITY.  
FT DISULFID 145 164 BY SIMILARITY.  
FT CARBOHYD 144 144 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CONFLICT 15 15 A -> G (IN REF. 2).  
SQ SEQUENCE 272 AA; 30153 MW; 06E7BB4156F0D08E CRC64;

Query Match 0.8%; Score 8; DB 1; Length 272;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCCPGH 623  
Db 123 CVPCCPGH 130

RESULT 4.

VBLL1\_TGMV  
ID VBLL1\_TGMV STANDARD; PRT; 293 AA.  
AC P03566;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE BL1 protein.  
GN BCL1.  
OS Tomato golden mosaic virus (TGMV).  
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.  
OX NCBI\_TaxID=10831;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA von Arnim A.G., Stanley J.;  
RL Submitted (XXY-1991) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP PRELIMINARY SEQUENCE FROM N.A.  
RA Hamilton W.D.O., Stein V.E., Coutts R.H.A., Buck K.W.;  
RT "Complete nucleotide sequence of the infectious cloned DNA components of tomato golden mosaic virus: potential coding regions and regulatory sequences."  
RT EMBO J. 3:2197-2205(1984).  
CC -!- SIMILARITY: BELONGS TO GEMINIVIRUSES BL1 PROTEIN FAMILY.  
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 173 ONWARD DUE TO A FRAMESHIFT ERROR.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M73794; AAA46584.1; -;  
EMBL; K02030; -; NOT\_ANNOTATED\_CDS.  
PIR; A04169; QOCVLG.  
DR InterPro; IPR000211; Gemini\_BL.  
DR Pfam; PF00845; Gemini\_BL1; 1.  
FT CONFLICT 11 11 V -> A (IN REF. 2).  
SQ SEQUENCE 293 AA; 32933 MW; E024E3F285C89012 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 293;  
Best Local Similarity 100.0%; Pred. No. 5.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 NYIESNRD 184  
Db 13 NYIESNRD 20

RESULT 5  
TR6B\_HUMAN  
ID TR6B\_HUMAN STANDARD; PRT; 300 AA.  
AC O95407;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Tumor necrosis factor receptor superfamily member 6B precursor (Decoy receptor for Fas ligand) (Decoy receptor 3) (Dcr3) (M68).  
GN TNFRSF6B OR DCR3 OR TR6.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Fetal lung;  
RX MEDLINE=99087326; PubMed=9872321;  
RA Pitti R.M., Marsters S.A., Lawrence D.A., Roy M., Kischkel F.C., Dowd P., Huang A., Donahue C.J., Sherwood S.W., Baldwin D.T., Godowski P.J., Wood W.I., Gurney A.L., Hillan K.J., Cohen R.L., Goddard A.D., Botstein D., Ashkenazi A.;  
RT "Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer."  
RT Nature 396:699-703(1998).  
RN [2]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-35.  
RC TISSUE=Prostate;  
RX MEDLINE=99253915; PubMed=10318773;  
RA Yu K.-Y., Kwon B., Ni J., Zhai Y.; Ebner R., Kwon B.S.;  
RT "A newly identified member of tumor necrosis factor receptor superfamily (TR6) suppresses LIGHT-mediated apoptosis."  
RT J. Biol. Chem. 274:13733-13736(1999).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=20122600; PubMed=10655513;  
RA Bai C., Connolly B., Metzker M.L., Hilliard C.A., Liu X., Sandig V., Soderman A., Galloway S.M., Liu Q., Austin C.P., Caskey C.T.;  
RT "Overexpression of M68/DCR3 in human gastrointestinal tract tumors independent of gene amplification and its location in a four-gene cluster."  
RT Proc. Natl. Acad. Sci. U.S.A. 97:1230-1235(2000).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Matthews L.;  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RA Strausberg R.;  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Decoy receptor for the cytotoxic ligands TNFS14/LIGHT and TNFSF6/FasL. Protects against apoptosis.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- TISSUE SPECIFICITY: Detected in fetal lung, brain and liver. Detected in adult stomach; spinal cord, lymph node, trachea, spleen, colon and lung. Highly expressed in several primary tumors from colon, stomach, rectum, esophagus and in SW480 colon carcinoma cells.  
CC -!- SIMILARITY: CONTAINS 4 TNFR-CYS REPEATS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF104419; AAD03056.1; -
CC EMBL; AF134240; AAD29688.1; -
CC EMBL; AF217796; AAF35244.1; -
CC EMBL; AF217793; AAF33685.1; -
CC EMBL; AF217794; AAF33686.1; -
CC EMBL; AL121845; CAC03668.1; -
CC EMBL; BC017065; AAH17065.1; -
CC Genew; HGNC:11921; TNFRSF6B.
CC MIM; 603361; -
CC HSSP; O14763; 1D0G.
CC InterPro; IPR001368; TNFR_c6.
CC Pfam; PF00020; TNFR_c6; 4.
CC ProDom; PD000771; TNFR_c6; 1.
CC SMART; SM00208; TNFR; 3.
CC PROSITE; PS00652; TNFR_NGFR_1; 2.
CC PROSITE; PS50050; TNFR_NGFR_2; 2.
CC Receptor; Apoptosis; Glycoprotein; Repeat; Signal.
KW SIGNAL 1 29
FT CHAIN 30 300 TUMOR NECROSIS FACTOR RECEPTOR
FT SUPERFAMILY MEMBER 6B.
FT REPEAT 31 70 TNFR-CYS 1.
FT REPEAT 72 113 TNFR-CYS 2.
FT REPEAT 115 150 TNFR-CYS 3.
FT REPEAT 152 193 TNFR-CYS 4.
FT DISULFID 49 62 BY SIMILARITY.
FT DISULFID 52 70 BY SIMILARITY.
FT DISULFID 73 88 BY SIMILARITY.
FT DISULFID 91 105 BY SIMILARITY.
FT DISULFID 95 113 BY SIMILARITY.
FT DISULFID 115 126 BY SIMILARITY.
FT DISULFID 132 150 BY SIMILARITY.
FT DISULFID 153 168 BY SIMILARITY.
FT DISULFID 174 193 BY SIMILARITY.
FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 300 AA; 32679 MW; F90AEE33718449AF CRC64;

Query Match 0.8%; Score 8; DB 1; Length 300;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 PCPPGTF 417
Db 152 PCPPGTF 159
|||||||

RESULT 6
DCP2_PEA STANDARD; PRT; 405 AA.
AC P51851;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update).
DE Pyruvate decarboxylase isozyme 2 (EC 4.1.1.1) (PDC) (Fragment).
GN PDC2.
OS Pism sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pism.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Miko;
RX MEDLINE=96215432; PubMed=8647075;
RA Muecke U., Wohlfarth T., Fiedler U., Baeumlein H.,
RA Ruecknagel K.P., Koenig S.;
RT "Pyruvate decarboxylase from Pism sativum. Properties, nucleotide
```

```
RT and amino acid sequences.";
RL Eur. J. Biochem. 237:373-382(1996).
CC -!- CATALYTIC ACTIVITY: A 2-oxo acid - an aldehyde + CO(2).
CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per
CC subunit.
CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z66544; CAA91445.1; -
CC HSSP; P06672; 1ZPD.
CC InterPro; IPR000399; TPP_enzyme.
CC Pfam; PF00205; TPP_enzymes; 1.
CC Pfam; PF02775; TPP_enzymes; 1.
CC PROSITE; PS00187; TPP_ENZYMES; PARTIAL.
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;
FT MULTIGENE FAMILY.
FT NON_TER 1
SQ SEQUENCE 405 AA; 44078 MW; 44DDCE90B38677FB CRC64;

Query Match 0.8%; Score 8; DB 1; Length 405;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLV 276
Db 50 KAVKPVLV 57
|||||||

RESULT 7
DCP1_TOBAC STANDARD; PRT; 418 AA.
ID DCP1_TOBAC
AC P51845;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyruvate decarboxylase isozyme 1 (EC 4.1.1.1) (PDC) (Fragment).
GN PDC1.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Samsun; TISSUE=Leaf;
RX MEDLINE=95375236; PubMed=7647304;
RA Bucher M., Brander K., Sbicego S., Mandel T., Kuhlmeier C.;
RT "Aerobic fermentation in tobacco pollen.";
RL Plant Mol. Biol. 28:739-750(1995).
CC -!- CATALYTIC ACTIVITY: A 2-oxo acid - an aldehyde + CO(2).
CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per
CC subunit.
CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).
CC -!- TISSUE SPECIFICITY: LEAVES.
CC -!- INDUCTION: ANAEROBICALLY.
CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```





```
RL Gene 275:31-37(2001).
CC -!- FUNCTION: Mediates activation of NF-kappa-B (By similarity). May
CC play a role in T-cell activation.
CC -!- SUBUNIT: Associates with TRAF1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -!- SIMILARITY: CONTAINS 1 TNFR-CYS REPEAT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB046039; BAB01621.1; -.
CC InterPro; IPR001368; TNFR_C6.
CC PROSITE; PS00652; TNFR_NGFR_1; FALSE_NEG.
CC PROSITE; PS50050; TNFR_NGFR_2; FALSE_NEG.
CC SMART; SM00208; TNFR; 1.
CC Receptor; Transmembrane; Glycoprotein; Signal.
CC SIGNAL 1 26
CC CHAIN 27 430
CC -----
CC DOMAIN 27 162
CC TRANSMEM 163 183
CC DOMAIN 184 430
CC REPEAT 50 90
CC DISULFID 51 65
CC DISULFID 71 90
CC CARBOHYD 149 149
CC SEQUENCE 430 AA; 45850 MW; BA8DE92593E1E859 CRC64;
CC -----
CC Query Match 0.8%; Score 8; DB 1; Length 430;
CC Best Local Similarity 100.0%; Pred. No. 7.5;
CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
QY 410 PCPPGTFS 417
Db 50 PCPPGTFS 57
CC -----
RESULT 10
DCP3_ORYSA STANDARD; PRT; 585 AA.
AC P51849;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyruvate decarboxylase isozyme 3 (EC 4.1.1.1) (PDC).
GN PDC3.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Indica-IR54; TISSUE=Callus;
RX MEDLINE=95148752; PubMed=7846174;
RA Hossain M.A., McGee J.D., Grover A., Dennis E., Peacock W.J.,
RA Hodges T.K.;
RA "Nucleotide sequence of a rice genomic pyruvate decarboxylase gene
RT that lacks introns: a pseudo-gene?";
RL Plant Physiol. 106:1697-1698(1994).
CC -!- CATALYTIC ACTIVITY: A 2-oxo acid = an aldehyde + CO(2).
CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per
CC subunit.
CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
QY 410 PCPPGTFS 417
Db 50 PCPPGTFS 57
CC -----
RESULT 10
DCP3_ORYSA STANDARD; PRT; 585 AA.
AC P51849;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyruvate decarboxylase isozyme 3 (EC 4.1.1.1) (PDC).
GN PDC3.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Indica-IR54; TISSUE=Callus;
RX MEDLINE=95148752; PubMed=7846174;
RA Hossain M.A., McGee J.D., Grover A., Dennis E., Peacock W.J.,
RA Hodges T.K.;
RA "Nucleotide sequence of a rice genomic pyruvate decarboxylase gene
RT that lacks introns: a pseudo-gene?";
RL Plant Physiol. 106:1697-1698(1994).
CC -!- CATALYTIC ACTIVITY: A 2-oxo acid = an aldehyde + CO(2).
CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per
CC subunit.
CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U07338; AAA68289.1; -.
CC HSP; P06169; IPVD.
CC InterPro; IPR000399; TPP_enzyme.
CC Pfam; PF00205; TPP_enzymes; 1.
CC Pfam; PF02775; TPP_enzymes_C; 1.
CC Pfam; PF02776; TPP_enzymes_N; 1.
CC PROSITE; PS00187; TPP_ENZYMES; FALSE_NEG.
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;
KW Multigene family.
FT ACT_SITE 71 BY SIMILARITY.
SQ SEQUENCE 585 AA; 62456 MW; F26AAFA5EC8AAE52 CRC64;
CC -----
CC Query Match 0.8%; Score 8; DB 1; Length 585;
CC Best Local Similarity 100.0%; Pred. No. 9.8;
CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
QY 269 KAVKPVLV 276
Db 231 KAVKPVLV 238
CC -----
RESULT 11
DCP2_ORYSA STANDARD; PRT; 603 AA.
ID DCP2_ORYSA STANDARD; PRT; 603 AA.
AC P51848;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyruvate decarboxylase isozyme 2 (EC 4.1.1.1) (PDC).
GN PDC2.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Indica-IR54; TISSUE=Callus;
RA Huq M.E., Hossain M.A., Hodges T.K.;
RT "Cloning and sequencing of a cDNA encoding pyruvate decarboxylase 2
RT gene from rice.";
RL (In) Plant Gene Register PGR95-072.
CC -!- CATALYTIC ACTIVITY: A 2-oxo acid = an aldehyde + CO(2).
CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per
CC subunit.
CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U27350; AAA90948.1; -.
CC EMBL; U38199; AAB40530.1; -.
CC HSP; P06169; IPVD.
CC InterPro; IPR000399; TPP_enzyme.
CC Pfam; PF00205; TPP_enzymes; 1.
CC Pfam; PF02775; TPP_enzymes_C; 1.
CC Pfam; PF02776; TPP_enzymes_N; 1.
CC PROSITE; PS00187; TPP_ENZYMES; FALSE_NEG.
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;
KW Multigene family.
FT ACT_SITE 89 BY SIMILARITY.
```



```
SQ SEQUENCE 603 AA; 64301 MW; 65B83A6468DCC2EB CRC64;

Query Match 0.8%; Score 8; DB 1; Length 603;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 269 KAVKPVLV 276
      |||||
Db 249 KAVKPVLV 256

RESULT 12
KRP2_RAT STANDARD; PRT; 671 AA.
AC Q62909;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Kinesin-related protein 2.
GN KRP2..
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Testis;
RX MEDLINE=96228687; PubMed=8688559;
RA Sperry A.O., Zhao L.-P.;
RT "Kinesin-related proteins in the mammalian testes: candidate motors
for meiosis and morphogenesis.";
RL Mol. Biol. Cell 7:289-305(1996).
RN [2]
RP REVISIONS.
RA Sperry A.O.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MOTOR PROTEIN, ACTIVE IN MEIOSIS.
CC -!- TISSUE SPECIFICITY: TESTIS. LOCALIZED TO THE MEIOTICALLY ACTIVE
CC CELLS OF THE SEMINIFEROUS EPITHELIA IN THE TESTIS.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U44979; AAC53528.1; -
HSP; P17119; 3KAR.
InterPro; IPR001752; kinesin_motor.
Pfam; PF00225; kinesin; 1.
PRINTS; PR00380; KINESINHEAVY.
SMART; SM00129; KISC; 1.
PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
Motor protein; Microtubules; ATP-binding; Coiled coil; Meiosis.
DOMAIN 1 198 GLOBULAR (POTENTIAL).
DOMAIN 199 540 KINESIN-MOTOR (BY SIMILARITY).
DOMAIN 541 671 COILED COIL (POTENTIAL).
NP_BIND 294 301 ATP (POTENTIAL).
SEQUENCE 671 AA; 75661 MW; F2B54598C78DE8DE CRC64;

Query Match 0.8%; Score 8; DB 1; Length 671;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 846 VISVPSKC 853
      |||||
Db 226 VISVPSKC 233
```

```
RESULT 13
MCAK_CRIGR STANDARD; PRT; 718 AA.
ID MCAK_CRIGR
AC P70096;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mitotic centromere-associated kinesin (MCAK) (Kinesin-like protein 6).
GN KNSL6.
OS Cricetulus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95122643; PubMed=7822426;
RA Wordeman L., Mitchison T.J.;
RT "Identification and partial characterization of mitotic centromere-
associated kinesin, a kinesin-related protein that associates with
centromeres during mitosis.";
RL J. Cell Biol. 128:95-105(1995).
RN [2]
RP REVISIONS.
RA Wordeman L.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PRESENT THROUGHOUT THE CELL CYCLE, ASSOCIATES WITH
CC CENTROMERES AT EARLY PROPHASE, AND REMAINS ASSOCIATED WITH THE
CC CENTROMERE UNTIL AFTER TELOPHASE.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U11790; AAB17358.2; -
HSP; P17119; 3KAR.
InterPro; IPR001752; kinesin_motor.
Pfam; PF00225; kinesin; 1.
PRINTS; PR00380; KINESINHEAVY.
SMART; SM00129; KISC; 1.
PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
Motor protein; Microtubules; ATP-binding; Coiled coil;
Nuclear protein.
DOMAIN 1 246 GLOBULAR (POTENTIAL).
DOMAIN 247 612 KINESIN-MOTOR (BY SIMILARITY).
DOMAIN 613 651 COILED COIL (POTENTIAL).
DOMAIN 689 716 COILED COIL (POTENTIAL).
NP_BIND 342 349 ATP (POTENTIAL).
DOMAIN 409 412 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SEQUENCE 718 AA; 80918 MW; 16ABD8BC66AD11B2.CRC64;

Query Match 0.8%; Score 8; DB 1; Length 718;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 846 VISVPSKC 853
      |||||
Db 274 VISVPSKC 281

RESULT 14
TAT_SIVCZ
ID TAT_SIVCZ STANDARD; PRT; 100 AA.
AC P17285;
```

DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE TAF protein (Transactivating regulatory protein).  
GN TAT.  
OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11723;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90259077; PubMed=2188136;  
RA Huet T., Cheynier R., Meyerhans A., Roelants G., Wain-Hobson S.;  
RT "Genetic organization of a chimpanzee lentivirus related to HIV-1.";  
RL Nature 345:356-359(1990).  
CC -!- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT ACTS BY BINDING TO THE  
CC TRANS-ACTIVATING RESPONSIVE SEQUENCE (TAR) RNA ELEMENT AND  
CC ACTIVATES TRANSCRIPTION INITIATION AND/OR ELONGATION FROM THE LTR  
CC PROMOTER.  
CC -!- SUBUNIT: BINDS CYCLIN T1 (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X52154; CAA36404.1; -.  
CC PIR; S09987; TNLJSI.  
CC HIV; X52154; TATSCPZ.  
CC InterPro; IPR001831; HIV\_Tat.  
CC Pfam; PF00539; Tat; 1.  
CC PRINTS; PR00055; HIVTATDOMAIN.  
KW Transcription regulation; Activator; RNA-binding; Nuclear protein;  
KW AIDS.  
SQ SEQUENCE 100 AA; 11209 MW; 1B78830B90EED50E CRC64;

Query Match 0.7%; Score 7; DB 1; Length 100;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 658 GSKNNQD 664 .  
Db 62 GSKNNQD 68  
|||||

RESULT 15  
RK24\_GUITH  
ID RK24\_GUITH STANDARD; PRT; 101 AA.  
AC O46905;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Chloroplast 50S ribosomal protein L24.  
GN RPL24.  
OS Guillardia theta (Cryptomonas phi).  
OG Chloroplast.  
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.  
OX NCBI\_TaxID=55529;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97283757; PubMed=9137835;  
RA Wang S.L., Liu X.-Q., Douglas S.E.;  
RT "The large ribosomal protein gene cluster of a cryptomonad plastid:  
RT gene organization, sequence and evolutionary implications.";  
RL Biochem. Mol. Biol. Int. 41:1035-1044(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99128221; PubMed=9929392;  
RA Douglas S.E., Penny S.L.;  
RT "The plastid genome of the cryptophyte alga, Guillardia theta;

RT complete sequence and conserved synteny groups confirm its common  
RT ancestry with red algae.";  
RL J. Mol. Evol. 48:236-244(1999).  
CC -!- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF041468; AAC35714.1; -.  
CC InterPro; IPR000302; KOW\_motif.  
CC InterPro; IPR003256; Ribosomal\_L24.  
CC Pfam; PF00467; KOW; 1.  
CC ProDom; PD001677; Ribosomal\_L24; 1.  
CC TIGRFAMS; TIGR01079; rplX\_bact; 1.  
CC PROSITE; PS01108; RIBOSOMAL\_L24; FALSE\_NEG.  
KW Ribosomal protein; Chloroplast.  
SQ SEQUENCE 101 AA; 11572 MW; 306E2E229C636131 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 101;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 792 LKNINIK 798  
Db 35 LKNINIK 41  
|||||

Search completed: May 12, 2003, 13:39:34  
Job time : 17.0658 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 07:48:38 ; Search time 46.4563 Seconds  
(without alignments)  
2762.174 Million cell updates/sec

Title: US-10-073-333A-2  
Perfect score: 5357  
Sequence: 1 MLFRARGPVRRGWRPAEA.....TCYFWKKKQKKKTILNLFN 963

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5357	100.0	963	22	TR16-short recepto
2	5307	99.1	1027	22	TR16-long receptor
3	2821.5	52.7	1013	21	Human CASB619 prot
4	2815.5	52.6	1001	22	Human TR13 recepto
5	2811.5	52.5	1013	22	Amino acid sequenc
6	2796	52.2	1013	22	Human PRO4985 poly
7	2712.5	50.6	911	22	Amino acid sequenc
8	2702.5	50.4	870	22	Amino acid sequenc
9	2526	47.2	464	22	Human SEC10 protei
10	2222	41.5	411	22	Human SEC5 protein

11	2000	37.3	750	22	AAB35328	Human TR13 recepto
12	1258.5	23.5	495	20	AAV59972	Human endometrium
13	1208.5	22.6	383	22	AAB83853	Amino acid sequenc
14	883	16.5	372	22	AAB85768	Human seven-transm
15	705.5	13.2	209	22	AAB83852	Amino acid sequenc
16	452	8.4	81	22	ABB39918	Peptide #7424 enco
17	452	8.4	81	22	ABB24471	Protein #6470 enco
18	452	8.4	81	22	AAM60663	Human brain expres
19	452	8.4	81	22	AAM73335	Human bone marrow
20	452	8.4	81	22	AAM33535	Peptide #7572 enco
21	452	8.4	81	23	ABG43186	Human peptide enco
22	444.5	8.3	208	21	AAB53442	Human colon cancer
23	442	8.3	74	22	AAV70281	Peptide #25. Unid
24	370.5	6.9	147	22	AAB83849	Peptide fragment o
25	353	6.6	60	22	ABB38686	Peptide #6192 enco
26	353	6.6	60	22	ABB23759	Protein #5758 enco
27	353	6.6	60	22	AAM59318	Human brain expres
28	353	6.6	60	22	AAV71867	Human bone marrow
29	353	6.6	60	22	AAM32149	Peptide #6186 enco
30	353	6.6	60	23	ABG41680	Human peptide enco
31	353	6.6	64	22	AAM72925	Human bone marrow
32	353	6.6	64	23	ABG42760	Human peptide enco
33	350	6.5	150	20	AAV12274	Human 5' EST secre
34	263	4.9	105	21	AAB26180	Human CASB619 prot
35	205	3.8	1605	21	AAB19805	Mouse laminin 2 ga
36	205	3.8	1605	21	AAB48454	Mouse laminin 8 po
37	205	3.8	1605	23	ABB81596	Mouse laminin 10 t
38	200	3.7	3635	23	ABB81589	Mouse laminin alph
39	200	3.7	3635	23	AAV50357	Mouse laminin-15 a
40	197.5	3.7	1572	21	AAB19806	Mouse laminin 2 ma
41	197.5	3.7	1572	21	AAB48455	Mouse laminin 8 po
42	197.5	3.7	1572	23	ABB81597	Mouse laminin 10 t
43	188.5	3.5	999	23	ABB80926	Human breast cance
44	188.5	3.5	999	23	ABG61893	Prostate cancer-as
45	182	3.4	2931	22	ABB68229	Drosophila melanog

ALIGNMENTS

RESULT 1  
AAB70255  
ID AAB70255 standard; protein; 963 AA.  
XX  
AC AAB70255;  
XX  
DT 10-MAY-2001 (first entry)  
XX  
DE TR16-short receptor protein.  
XX  
KW TR16 recptor; tumour necrosis factor receptor superfamily;  
KW apoptosis; inflammatory; cancer; immune; neurodegenerative.  
XX  
OS Unidentified.  
XX  
PN WO200112671-A1.  
XX  
PD 22-FEB-2001.  
XX  
PF 10-AUG-2000; 2000WO-US21885.  
XX  
PR 12-AUG-1999; 99US-0148348.  
PR 13-AUG-1999; 99US-0148683.  
PR 13-AUG-1999; 99US-0148870.  
PR 16-AUG-1999; 99US-0148758.  
PR 17-AUG-1999; 99US-0149181.  
PR 18-AUG-1999; 99US-0149453.  
PR 19-AUG-1999; 99US-0149498.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Young PE, Baker KP;  
XX

DR WPI; 2001-138754/14.

XX New nucleic acid molecule encoding a TR16 tumor necrosis factor

PT receptor polypeptide, useful for the diagnosis and treatment of cancer,

PT autoimmune disorders and cardiovascular diseases -

XX Claim 1; Fig 1; 286pp; English.

PS The present invention relates to a TR16 receptor (tumour necrosis

XX factor receptor superfamily). The invention is useful treating

CC diseases and disorders associated with the inhibited or increased

CC apoptosis. In particular inflammatory diseases, cancers, immune and

CC neurodegenerative disorders may be treated.

XX Sequence 963 AA;

SQ Query Match 100.0%; Score 5357; DB 22; Length 963;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFRARGPVRGRGWGRPAEAPRRGRSPWPSPAWICCWALAGCQAAWAGDLPSSSSRPLPP 60

DB 1 MLFRARGPVRGRGWGRPAEAPRRGRSPWPSPAWICCWALAGCQAAWAGDLPSSSSRPLPP 60

QY 61 CQEKDYHFYEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120

DB 61 CQEKDYHFYEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120

QY 121 SKCGEGYSLGSGIKFDEWDELPAFNSIATFMDTVVGPSRDPGCGNNSWIIPRGNYIE 180

DB 121 SKCGEGYSLGSGIKFDEWDELPAFNSIATFMDTVVGPSRDPGCGNNSWIIPRGNYIE 180

QY 181 SNRDDCTVSLIYAVHLKSGYVFEFYQYVDNNIFFEYQNDQCQEMDTTTTDKWVKLTDN 240

DB 181 SNRDDCTVSLIYAVHLKSGYVFEFYQYVDNNIFFEYQNDQCQEMDTTTTDKWVKLTDN 240

QY 241 GEWGSYMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCPKPGTFSN 300

DB 241 GEWGSYMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCPKPGTFSN 300

QY 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360

DB 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360

QY 361 KQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPNPGFYNNGSSSCHPCPPGTFSDGT 420

DB 361 KQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPNPGFYNNGSSSCHPCPPGTFSDGT 420

QY 421 KECRPCAGTEPALGFYKWNVLPGNMKTSFCNVGNSKCDGMNGWEVAGDHIQSGAGGS 480

DB 421 KECRPCAGTEPALGFYKWNVLPGNMKTSFCNVGNSKCDGMNGWEVAGDHIQSGAGGS 480

QY 481 DNDYLIILNLHIPGFKPPTSMGTAGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540

DB 481 DNDYLIILNLHIPGFKPPTSMGTAGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540

QY 541 SWGGTKEQAYTHIIFKNATFTTWFQRTNQODNRRFINDMVKIYSITATNAVGVAS 600

DB 541 SWGGTKEQAYTHIIFKNATFTTWFQRTNQODNRRFINDMVKIYSITATNAVGVAS 600

QY 601 SCRACALCSEQSGSSCVPCPPGHYIEKETNQCKECPDPTYLSIHQVYGKEACIPCGPGSK 660

DB 601 SCRACALCSEQSGSSCVPCPPGHYIEKETNQCKECPDPTYLSIHQVYGKEACIPCGPGSK 660

QY 661 NNQDHSVYSDCFFYHEKENQILHYDFSNLSSVGLMNGPSTSKGTKYFHFNLSLCGH 720

DB 661 NNQDHSVYSDCFFYHEKENQILHYDFSNLSSVGLMNGPSTSKGTKYFHFNLSLCGH 720

QY 721 EGKKMALCTNNITDFTVKEIVAGSDDTNLVGAFCVQCSTIIPSESKGFRAALSSQSIILA 780

DB 721 EGKKMALCTNNITDFTVKEIVAGSDDTNLVGAFCVQCSTIIPSESKGFRAALSSQSIILA 780

QY 781 DTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSSTATTSINGRSTAVKMRNCNPT 840

DB 781 DTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSSTATTSINGRSTAVKMRNCNPT 840

QY 841 KSGAGVISVPSKCPAGTCDCGCTFYFLWESAEACPLCTEHDHFEIEGACKRGFQETLYVWN 900

DB 841 KSGAGVISVPSKCPAGTCDCGCTFYFLWESAEACPLCTEHDHFEIEGACKRGFQETLYVWN 900

QY 901 EPKWKIKGISLPEKKLATCETVDFWLKVGAGVGAFATVLLVALTCYFWKKNQKKKKTILN 960

DB 901 EPKWKIKGISLPEKKLATCETVDFWLKVGAGVGAFATVLLVALTCYFWKKNQKKKKTILN 960

QY 961 LFN 963

DB 961 LFN 963

RESULT 2

AAB70256

ID AAB70256 standard; protein; 1027 AA.

XX

AC AAB70256;

XX

DT 10-MAY-2001 (first entry)

XX

DE TR16-long receptor protein.

XX

KW TR16-receptor; tumour necrosis factor receptor superfamily;

KW apoptosis; inflammatory; cancer; immune; neurodegenerative.

XX

OS Unidentified.

XX

PN WO200112671-A1.

XX

PD 22-FEB-2001.

XX

PF 10-AUG-2000; 2000WO-US21885.

XX

PR 12-AUG-1999; 99US-0148348.

PR 13-AUG-1999; 99US-0148683.

PR 13-AUG-1999; 99US-0148870.

PR 16-AUG-1999; 99US-0148758.

PR 17-AUG-1999; 99US-0149181.

PR 18-AUG-1999; 99US-0149453.

PR 19-AUG-1999; 99US-0149498.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Ruben SM, Young PE, Baker KP;

XX

DR WPI; 2001-138754/14.

XX

PT New nucleic acid molecule encoding a TR16 tumor necrosis factor

PT receptor polypeptide, useful for the diagnosis and treatment of cancer,

PT autoimmune disorders and cardiovascular diseases -

XX

PS Disclosure; Fig 4; 286pp; English.

XX

CC The present invention relates to a TR16 receptor (tumour necrosis

CC factor receptor superfamily). The invention is useful treating

CC diseases and disorders associated with the inhibited or increased

CC apoptosis. In particular inflammatory diseases, cancers, immune and

CC neurodegenerative disorders may be treated.

XX

SQ Sequence 1027 AA;

Query Match 99.1%; Score 5307; DB 22; Length 1027;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 953; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFRARGPVRGRGWGRPAEAPRRGRSPWPSPAWICCWALAGCQAAWAGDLPSSSSRPLPP 60

DB 1 MLFRARGPVRGRGWGRPAEAPRRGRSPWPSPAWICCWALAGCQAAWAGDLPSSSSRPLPP 60



QY 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120  
Db 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120  
QY 121 SKCGEGTYSGLSGIKFDEWDELPAFGSNIATFMDTVVGPDSRDPGCGNNSWIPRGNYYIE 180  
Db 121 SKCGEGTYSGLSGIKFDEWDELPAFGSNIATFMDTVVGPDSRDPGCGNNSWIPRGNYYIE 180  
QY 181 SNRDDCTVSLIYAVHLKSGYVFEFYQYVDNNIFFEFFIIONDQCQEMDTTDDKWVKLTDN 240  
Db 181 SNRDDCTVSLIYAVHLKSGYVFEFYQYVDNNIFFEFFIIONDQCQEMDTTDDKWVKLTDN 240  
QY 241 GEWGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECPKPGTFSN 300  
Db 241 GEWGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECPKPGTFSN 300  
QY 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360  
Db 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360  
QY 361 KTOIMYKWKIEPKICREDLTDAIRLPPSGEKKDCPPCNPFGFYNNGSSSCHPCPPGTFSDGT 420  
Db 361 KTOIMYKWKIEPKICREDLTDAIRLPPSGEKKDCPPCNPFGFYNNGSSSCHPCPPGTFSDGT 420  
QY 421 KECRPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480  
Db 421 KECRPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480  
QY 481 DNDYLILNLHIPGPKPPTSMTGATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540  
Db 481 DNDYLILNLHIPGPKPPTSMTGATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540  
QY 541 SWGGTKEKQAYTHIIFKNATFTTFAFORNQGDNRREINDMVKIYSTATNAVGVAS 600  
Db 541 SWGGTKEKQAYTHIIFKNATFTTFAFORNQGDNRREINDMVKIYSTATNAVGVAS 600  
QY 601 SCRACALGSEQSGSSCVPCPPGGHYIEKETNQCKECPDPTLYLSIHQVYGEACIPCGPGSK 660  
Db 601 SCRACALGSEQSGSSCVPCPPGGHYIEKETNQCKECPDPTLYLSIHQVYGEACIPCGPGSK 660  
QY 661 NNQDHSVYCSDCFYHEKENQILHYDFSLSLVGSLMNGPSFTSKGTGYFHFNFNISLCGH 720  
Db 661 NNQDHSVYCSDCFYHEKENQILHYDFSLSLVGSLMNGPSFTSKGTGYFHFNFNISLCGH 720  
QY 721 EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFCVQSTIIPSESKGFRAALSSQSIIILA 780  
Db 721 EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFCVQSTIIPSESKGFRAALSSQSIIILA 780  
QY 781 DTFIGTVETTLKININIKEDMFPVPPTSQIPDVHFFYKSTATTSCINGRSTAVKMRCNPT 840  
Db 781 DTFIGTVETTLKININIKEDMFPVPPTSQIPDVHFFYKSTATTSCINGRSTAVKMRCNPT 840  
QY 841 KSGAGVISVPKCPAGTCDGCTFYFLWESAEACPLCTEHDFHEIEGACKRGFQETLYVWN 900  
Db 841 KSGAGVISVPKCPAGTCDGCTFYFLWESAEACPLCTEHDFHEIEGACKRGFQETLYVWN 900  
QY 901 EPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAFNAVLLVALTCYFWKKNQK 953  
Db 901 EPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAFNAVLLVALTCYFWKKNQK 953

RESULT 3  
AAB26179  
ID AAB26179 standard; Protein; 1013 AA.

XX AAB26179;  
AC AAB26179;  
XX 12-FEB-2001 (first entry)  
DT Human CASB619 protein #1.

XX Human; CASB619; cancer; autoimmune disease; immunogen; vaccine;  
KW epitope.

XX Homo sapiens.  
OS WO200058460-A2.  
XX 05-OCT-2000.  
XX 20-MAR-2000; 2000WO-EP02478.  
XX 26-MAR-1999; 99GB-0007113.  
PR 25-SEP-1999; 99GB-0022858.  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX Bruck CEM, Cassart J, Coche T, Vinals De Bassols YC;  
PI WPI; 2000-664923/64.  
XX N-PSDB; AAA95442.  
PT Novel CASB619 polypeptides useful for diagnosing, and as vaccines for  
PT prophylactic and therapeutic treatment of, cancers, particularly  
PT ovarian and colon carcinoma, and autoimmune diseases -  
XX Claim 4; Page 54-56; 68pp; English.  
PS The present sequence comprises the human CASB619 protein sequence. This  
XX protein is thought to be specifically or over-expressed in tumour cells,  
CC and so can be used as a target for antigen-specific immune responses  
CC which can cause destruction of the tumour cell. In addition, the protein  
CC and gene can be used in cancer diagnosis, in the treatment of autoimmune  
CC diseases and in vaccines against cancer and autoimmune disease. The  
CC invention provides a number of epitopes derived from the protein which  
CC can be used as immunogens.  
XX Sequence 1013 AA;

Query Match 52.7%; Score 2821.5; DB 21; Length 1013;  
Best Local Similarity 53.0%; Pred. NO. 3.8e-214;  
Matches 502; Conservative 161; Mismatches 260; Indels 25; Gaps 12;

QY 23 RGRSPWPSPAWICWALAGCAAAG---DLPSSSSRPLPPCQEKDYHFEYTECDSSGSR 79  
Db 15 RGRTERIPR---LWRL---LLWAGTAFQVTOGTGPELHACKESEYHYEYACDSTGSR 67  
QY 80 WRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVCSCGEGTYSGLSGIKFDEW 139  
Db 68 WRVAVPHTPGLCTSLPDPVKGTGTECSFCSCNAGEFLDMKDQCKPCAEGRYSLTGTGIREDEW 127  
QY 140 DELPAGFSNIATFMDTVVGPDSRDPGCGNNSSWIPRGNYYIESNRDDCTVSLIYAVHLKKS 199  
Db 128 DELPHGFEASLSANMELDDSAAES-TGNCTSSKWVPRGDYIASNTDECTATLMYAVNLKQS 186  
QY 200 GYVFFEYQYVDNNIFFEFFIIONDQCQEMDTTDDKWVKLTDNGEWSHVMKSGTNILYW 259  
Db 187 GTVNFYYYPDSSIIFEFFVQNDQCQP-NABDSRWMTTEKG-WEFHSHVELNRGNVLYW 244  
QY 260 RTTGILMGSKAVKPVLVKNITIEGVAYTSECPKPGTFSNKPGEFNCQVCPRNTYSEKG 319  
Db 245 RTTAFSVWTKVPKPVLVNRNIAITGVAYTSECPKPGTYADKQGSFCKLCPANSYSNKG 304  
QY 320 AKECIRCKDDSQFS-GSSECTERPPCTTKDYFQIHTPCDEEGKTOIMYKWKIEPKICRED 377  
Db 305 ETSCHQC-DPDKYSEKSGSSCNVRPACTDKDYFYTHTACDANGETQLMYKWKAPKICSED 363  
QY 378 LTDAIRLPPSGEKKDCPPCNPFGFYNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFE 437  
Db 364 LEGAVKLPAASGVKTHCPCPNPGFFKTNNSTCQPCPYGSYNGS-DCTRCPAGTEPAVGFE 422  
QY 438 YKWWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSNDYLILNLHIPGFKPP 497  
Db 423 YKWWNTLPTNMETTVLSGINFEYKGMTGWEVAGDHIYTAAGASDNDFMILTLLVVPGRFP 482  
QY 498 TS-MTGATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVESWGCTKEKQAYTHIIF 556

Db 483 QSMADTENKEVARITFVETLCSVNCLELYFMVGVNSRTNTPVETWKGSKGQSYTYIE 542

Qy 557 KNATFTTFAFORTNQDNRFFINDMKIYSITATNAVGVASSCRACALGSEQSGSSC 616

Db 543 ENTTSFTWAFORTTFHEASRKYTNDVAKIYSINVTVMNGVASYCRPCALEASDVGSSC 602

Qy 617 VPCPPGHIYIEKETNOCKECPDPTLSIHQVYGKEACIPCGPGSKNNQDHSVYSCDFFYH 676

Db 603 TSCPAGYYIDRDSGTCHSCPNTILKAHQPYGVQACVPCGPGTKNNKIHSCLYNDCTFSR 662

Qy 677 EKENQILHYDFSLSVGLMNGPSFTSGTKYFHHFFNISLCGHEGKKMALCTNNITDFT 736

Db 663 NTPTRTFNYNFSALANTVTLAGGSPFTSKGLKYFHHFTLSLCGNQGRKMSVCTDNVTDLR 722

Qy 737 VKEIVAGSDDYTNLVGAFVCSQTIIPSESKGFRALSSQSIILADTFIGVTVETTLKNIN 796

Db 723 IPE---GESGFSKSITAYVQAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGIT 779

Qy 797 IKEDMFPVPTSQIPDVHFFYKSSSTATTSINGRSTAVKMRNPTKSGAGVISVPSKCPAG 856

Db 780 SPAELFHLESGLIPDVIFFYRSNDVTQSCSSGRSTIRVRCSPQKTVPGSLLLPCTCSDG 839

Qy 857 TCDGCTFYFLWESAEACPLCTEHDHFEIEGACKRGFOETLYVWNEPKWCICKISLPEKKL 916

Db 840 TCDGCNHFHFLWESAAACPLCSVADYHAIIVSSVAGIQKTTYVWREPKLCSGISLPEQRY 899

Qy 917 ATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKNQKK---KTILN 960

Db 900 TICKTIDFWLKVGISAGTCTAILLTVLTCYFWKKNQKLEYKYSKLVMN 947

RESULT 4

AAB35333

ID AAB35333 standard; Protein; 1001 AA.

XX AAB35333;

AC AAB35333;

XX AAB35333;

DT 08-MAY-2001 (first entry)

XX Human TR13 receptor protein SEQ ID NO: 40.

DE Human; tumor necrosis factor receptor; TR13; TR14; infection;

XX cancer; autoimmune disease; allergy; inflammatory disease;

KW graft rejection; apoptosis; cardiovascular disease; aneurysm.

XX Homo sapiens.

OS WO200105834-A1.

XX 25-JAN-2001.

XX 14-JUL-2000; 2000WO-US19343.

XX 16-JUL-1999; 99US-0144087.

XX 18-AUG-1999; 99US-0149450.

XX 20-AUG-1999; 99US-0149712.

XX 10-SEP-1999; 99US-0153089.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Ruben SM, Ni J, Young PE;

XX WPI: 2001-112682/12.

DR Nucleic acids encoding 2 human tumor necrosis factor receptor

XX polypeptides ((TR13) and (TR14)), useful for the prevention, diagnosis

PT and treatment of, e.g. cancers, acquired immune deficiency syndrome and

PT hypohidrotic ectodermal dysplasia -

XX Claim 40; Page 398-401; 418pp; English.

XX The present invention provides the protein and coding sequences of the

CC

CC human tumour necrosis factor receptors TR13 and TR14. These sequences are

CC useful in the diagnosis and treatment of many diseases, including cancer,

CC autoimmune diseases, cardiovascular disorders, allergies,

CC neurodegenerative diseases, graft rejection, inflammation; aneurysms and

CC infections.

XX Sequence 1001 AA;

SQ

Query Match 52.6%; Score 2815.5; DB 22; Length 1001;

Best Local Similarity 52.8%; pred. No. 1.1e-213;

Matches 501; Conservative 161; Mismatches 261; Indels 25; Gaps 12;

Qy 23 RGRSPWPSPAWICCWALAGCAAWAG---DLPSSSSRPLPPCQEKDYHFEYTECDSSGR 79

Db 15 RGRTERIPR---LWRL---LLWAGTAFQVTOGTGPELHACKESEYHYEYACDSTCSR 67

Qy 80 WRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVCSKCGEGTYSLSGSGIKFDEW 139

Db 68 WRVAVPHTPGLCTSLPDPVKGTCSFSCNAGEFLDMKDQCKPACAEGRYSLGTGIRFDEW 127

Qy 140 DELPACFSNIATFMDTVVGPDSRDPDGCNNSSWIPRGNVIESNRDDCTVSLIYAVHLKKS 199

Db 128 DELPHGFASLSANMELDDSAES-TGNCSTSSKVVPRGDYAFNTDECTATLMYAVNKLKS 186

Qy 200 GYVFEFYQYVDNNIFFEEFIQNDQCEMDTTTDKWKVLTNDGEWSHSMVKSGTNNILYW 259

Db 187 GTVNFEEYYPDSSIIFFEFVQNDQCP-NADDSRWMTTEKG-WEFHSVELNRGNVLYW 244

Qy 260 RTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCPKPGTFNSKPGSFNCQVCPRNTYSEK 319

Db 245 RTTAFSVMTKVPKPVLVRNIATGVAYTSECFPCPKPGTYADQGSFCKLCPANSYSNKG 304

Qy 320 AKECIRCKDDSQFS--GSSECTERPPCTTKDYFQIHTPCDEGKTQIMYKWIPEKICRED 377

Db 305 ETSCHQC-DPDKYSEKSSSCNVPRACTIDKDYFYHTACDANGETQLMYKWAKPKICSED 363

Qy 378 LTDAIRLPPSGEKKDCPPCPNPGFYNNSSSCHPPPGTFSGTEKRCPCPAGTEPALGFE 437

Db 364 LEGAVKLPASGVKTHCPCNPGFEKTNNSTCQPCPYGSYNGS-DCTRCPAGTEPAVGFE 422

Qy 438 YKWNVLPNGMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHHPGFKPP 497

Db 423 YKWNVLPNGMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHHPGFKPP 482

Qy 498 TS-MTGATGSELGRITFVFETLCSADCVLFMVVDINRKSTNVVSNESGTEKQAYTHIIF 556

Db 483 QSVMDATENKEVARITFVETLCSVNCLELYFMVGVNSRTNTPVETWKGSKGQSYTYIE 542

Qy 557 KNATFTTFAFORTNQDNRFFINDMKIYSITATNAVGVASSCRACALGSEQSGSSC 616

Db 543 ENTTSFTWAFORTTFHEASRKYTNDVAKIYSINVTVMNGVASYCRPCALEASDVGSSC 602

Qy 617 VPCPPGHIYIEKETNOCKECPDPTLSIHQVYGKEACIPCGPGSKNNQDHSVYSCDFFYH 676

Db 603 TSCPAGYYIDRDSGTCHSCPNTILKAHQPYGVQACVPCGPGTKNNKIHSCLYNDCTFSR 662

Qy 677 EKENQILHYDFSLSVGLMNGPSFTSGTKYFHHFFNISLCGHEGKKMALCTNNITDFT 736

Db 663 NTPTRTFNYNFSALANTVTLAGGSPFTSKGLKYFHHFTLSLCGNQGRKMSVCTDNVTDLR 722

Qy 737 VKEIVAGSDDYTNLVGAFVCSQTIIPSESKGFRALSSQSIILADTFIGVTVETTLKNIN 796

Db 723 IPE---GESGFSKSITAYVQAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGIT 779

Qy 797 IKEDMFPVPTSQIPDVHFFYKSSSTATTSINGRSTAVKMRNPTKSGAGVISVPSKCPAG 856

Db 780 SPAELFHLESGLIPDVIFFYRSNDVTQSCSSGRSTIRVRCSPQKTVPGSLLLPCTCSDG 839

Qy 857 TCDGCTFYFLWESAEACPLCTEHDHFEIEGACKRGFOETLYVWNEPKWCICKISLPEKKL 916

Db 840 TCDGCNHFHFLWESAAACPLCSVADYHAIIVSSVAGIQKTTYVWREPKLCSGISLPEQRY 899

Qy 917 ATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKNQKK---KTILN 960





KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;  
KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;  
KW adipocyte; A-peptide; factor VIIA; gene therapy.

XX  
OS Homo sapiens.

XX WO200140466-A2.

XX  
PD 07-JUN-2001.

XX  
PF 01-DEC-2000; 2000WO-US32678.

XX  
PR 01-DEC-1999; 99WO-US28301.

PR 01-DEC-1999; 99WO-US28634.

PR 02-DEC-1999; 99WO-US28551.

PR 02-DEC-1999; 99WO-US28564.

PR 02-DEC-1999; 99WO-US28565.

PR 09-DEC-1999; 99US-0170262.

PR 16-DEC-1999; 99WO-US30095.

PR 20-DEC-1999; 99WO-US30911.

PR 20-DEC-1999; 99WO-US30999.

PR 30-DEC-1999; 99WO-US31243.

PR 06-JAN-2000; 2000WO-US00277.

PR 06-JAN-2000; 2000WO-US00376.

PR 11-FEB-2000; 2000WO-US03565.

PR 18-FEB-2000; 2000WO-US04341.

PR 18-FEB-2000; 2000WO-US04342.

PR 22-FEB-2000; 2000WO-US04414.

PR 24-FEB-2000; 2000WO-US04914.

PR 24-FEB-2000; 2000WO-US05004.

PR 01-MAR-2000; 2000WO-US05601.

PR 20-MAR-2000; 2000WO-US07377.

PR 21-MAR-2000; 2000WO-US07532.

PR 30-MAR-2000; 2000WO-US08439.

PR 17-MAY-2000; 2000WO-US13705.

PR 22-MAY-2000; 2000WO-US14042.

PR 30-MAY-2000; 2000WO-US14941.

PR 02-JUN-2000; 2000WO-US15264.

PR 10-NOV-2000; 2000WO-US30873.

XX  
PA (GETH ) GENENTECH INC.

PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;  
PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;  
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;

XX  
DR WPI; 2001-408281/43.  
DR N-PSDB; AAS21262.

XX  
PT Isolated, secretory and transmembrane PRO polypeptide used to detect  
PT other PRO polypeptides, link bioactive molecules to cells expressing  
PT PRO polypeptides, and detect the presence of mammalian tumours e.g.  
PT lung, breast, prostate, cervical .

XX  
PS Claim 12; Fig 38; 813pp; English.

XX  
CC AAU12172-AAU12446 represent novel human secretory and transmembrane  
CC PRO polypeptides. The PRO polypeptides are useful to detect other  
CC PRO polypeptides, to link bioactive molecules to cells expressing  
CC PRO polypeptides, to modulate biological activities of cells expressing  
CC PRO polypeptides, and to detect the presence of mammalian lung, colon,  
CC breast, prostate, rectal, cervical or liver tumours by comparing PRO  
CC polypeptide expression in a cell sample to that in a control sample.  
CC Some of the 275 sequences are also useful to stimulate the release of  
CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the  
CC proliferation or differentiation of chondrocytes, the proliferation or  
CC gene expression in pericyte cells, the release of proteoglycans from  
CC cartilage, the proliferation of inner ear utricular supporting cells or  
CC of T-lymphocytes, the release of a cytokine from peripheral blood  
CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of  
CC the PRO polypeptides may modulate glucose or free fatty acid uptake by  
CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide  
CC to factor VIIA. The PRO polypeptides can be used in assays to identify



<b>RESULT 7</b>					
AAB83850					
ID	AAB83850	standard; Protein;	911 AA.		
XX	AC				
XX	AC				
XX	AA	B83850;			
DT	23-JUL-2001	(first entry)			
XX					
DE	Amino acid sequence of a human protein expressed in tumour cells.				
XX					
KW	Tumour cell; immunological disease; autoimmune disease; cancer;				
KW	infection.				
XX					
OS	Homo sapiens.				
XX					
PX	WO200131003-A1.				
PD					
XX	03-MAY-2001.				
PF	30-OCT-2000;	2000WO-FR03032.			
XX					
PR	29-OCT-1999;	99FR-0013629.			
XX					
PA	(FABR )	FABRE MEDICAMENT SA PIERRE.			
XX					
PI	Delneste Y,	Magistrelli G,	Jeannin P,	Bonnefoy J;	
XX					
DR	WPI;	2001-328651/34.			
DR	N-PSDB;	AAF89774.			
XX					
PT	New nucleic acid, expressed in tumours and lymphoid tissue is useful for identifying agents for treating tumours and autoimmune disease -				
XX					
PS	Claim 10;	Page 60-63;	85pp;	French.	
XX					
CC	The present sequence represents a human protein expressed in tumour cells. The polynucleotide is useful for screening cDNA/genomic DNA banks and for cloning isolated DNA; identifying mutant forms of the gene that encodes a human protein, where the mutations are associated with abnormal gene expression, or promoters and regulators of the gene, particularly for diagnosis; for recombinant expression of the derived protein; as probes and primers for detection and amplification; and as antisense therapeutics. The tumour expressed protein is useful for raising specific antibodies and to screen agents that modulate its activity, bind to it or interact with it. These agents are potentially useful for treatment or prevention of diseases associated with abnormal expression/activity of the protein, particularly immunological diseases (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic infections.				
XX					
SQ	Sequence	911 AA;			
Query Match					
Best Local Similarity 50.6%; Score 2712.5; DB 22; Length 911;					
Matches 482; Conservative 156; Mismatches 253; Indels 21; Gaps 11					
QY	23	RGRSPWPSPAWICCWALAGCQAAG--DLPSSSRPLPQCQEKDYHFEYTECDSSGSR	79		
Db	15	RGRTERRIPR--LWRL---LLWAGTAFOVTQTGTGPELHACKSEYHYEYTACDSTGSR	67		
QY	80	WRVAIPNSAVDCSLDPVRGKECTFCASGEYLEMKNQVCKCGEGTYSLGSGIKFDEW	139		
Db	68	WRVAVPHTPGLCTSLPDPVKGTSCFSNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEW	127		
QY	140	DELPAFGSNIATFMDTVVGPSDSRPDGCNNSSWIIPRGNYIESNRDDCTVSLIYAVHLKKS	199		
Db	128	DELPHGFASLSANMELDDSAES-TGNCTSSKWVPRGDYIASNTDECTATLMYAVNLKQS	186		
QY	200	GYVFFEYQYVDNNIFFEFFIQNDQCQEMDTTTDKWKVLTDNGEWGSHSVMLKSGTNILYW	259		
Db	187	GTVNFYYYYPDSSIIEFFVQNDQCQP-NADDSRWMMKTTEKG-WEFHSVELNRGNVLYW	244		



XX SQ Sequence 464 AA; Query Match 47.2%; Score 2526; DB 22; Length 464; Best Local Similarity 98.9%; Pred. No. 3.1e-191; Matches 452; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

QY 115 MKNQVCSKCGEGTYSLSGSIKFDEWDELPAAGFSNIATFMDTVVGPDSRDPDGNSSSWIP 174  
Db 1 MKNQVCSKCGEGTYSLSGSIKFDEWDELPAAGFSNIATFMDTVVGPDSRDPDGNSSSWIP 60

QY 175 RGNYESNRDDCTVSLIYAVHLKSGYVFFEQYVDNFIFFEQNDQCQEMDITTDKW 234  
Db 61 RGNYESNRDDCTVSLIYAVHLKSGYVFFEQYVDNFIFFEQNDQCQEMDITTDKW 120

QY 235 VKLTDNGEGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCK 294  
Db 121 VKLTDNGEGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCK 180

QY 295 PGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTERPPCTTKDYFQI 352  
Db 181 PGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSSEGSSECTERPPCTTKDYFQI 240

QY 353 HTPCDEEGKTQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPFGYNNGSSSCHPCP. 412  
Db 241 HTPCDEEGKTQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPFGYNNGSSSCHPCP 300

QY 413 PGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 472  
Db 301 PGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 360

QY 473 IQSGAGGSDNDYLILNLHIPGFKPPTSMGTATGSELGRITFTFETLCSADCVLVFMVDIN 532  
Db 361 IQSGAGGSDNDYLILNLHIPGFKPPTSMGTATGSELGRITFTFETLCSADCVLVFMVDIN 420

QY 533 RKSTNVVESWGTKEKQAYTHIIFKNATFTFTWAFQR 569  
Db 421 RKSTNVVESWGTKEKQAYTHIIFKNATFTFTWGIPR 457.

RESULT 10  
AAB48372  
ID AAB48372 standard; Protein; 411 AA.  
XX , AC AAB48372;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE Human SEC5 protein sequence (clone ID 1795045.0.61).

XX  
KW SECX; cytostatic; gynecological; gene therapy; screening assay; human;  
KW SEC5; chromosomal mapping; forensic biology; cell proliferation; cancer;  
KW cell differentiation; immune associated disorder; gestational disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200078802-A2.  
XX  
PD 28-DEC-2000.  
XX  
PF 23-JUN-2000; 2000WO-US17328.  
XX  
PR 23-JUN-1999; 99US-0140584.  
PR 20-JUL-1999; 99US-0144722.  
PR 16-SEP-1999; 99US-0154520.  
PR 22-JUN-2000; 2000US-0604286.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shimkets RA, Fernandes E, Vernet C, Yang M, Boldog FL;  
PI Herrmann JL;  
XX  
DR WPI; 2001-071385/08.

DR N-PSDB; AAC84886.  
XX Polynucleotides encoding SECX proteins useful for treating disease  
PT characterized by an aberrant level of cell proliferation and/or  
PT differentiation like cancer or immune associated disorders -  
XX Claim 1; Fig 6; 132pp; English.  
PS  
XX The invention relates to human SECX polypeptides and polynucleotides  
CC encoding them. The SECX polypeptides can be expressed by standard  
CC recombinant methodology. The SECX polypeptides are useful for treating  
CC or preventing a SECX-associated disorder. The invention is useful in  
CC screening assays; detection assays (e.g. chromosomal mapping, cell and  
CC tissue typing, forensic biology); predictive medicine (diagnostic assays,  
CC prognostic assays, monitoring clinical trials, and pharmacogenomics); and  
CC methods of treatment (e.g. therapeutic and prophylactic), especially  
CC disorders characterized by aberrant cell proliferation and/or  
CC differentiation like cancer or immune associated disorders or gestational  
CC disease. The present sequence represents a SEC5 protein.  
XX  
SQ Sequence 411 AA;  
Query Match 41.5%; Score 2222; DB 22; Length 411;  
Best Local Similarity 98.5%; Pred. No. 2.9e-167;  
Matches 397; Conservative 1; Mismatches 3; Indels 2; Gaps 1;

QY 169 NSSWIPRGNYESNRDDCTVSLIYAVHLKSGYVFFEQYVDNFIFFEQNDQCQEMD 228  
Db 2 HSSWIPRGNYESNRDDCTVSLIYAVHLKSGYVFFEQYVDNFIFFEQNDQCQEMD 61

QY 229 TTIDKWKVLTNDNGEGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTS 288  
Db 62 TTIDKWKVLTNDNGEGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTS 121

QY 289 ECFPCPKPGTFSNKPFSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTERPPCTT 346  
Db 122 ECFPCPKPGTFSNKPFSFNCQVCPRNTYSEKGAKECIRCKDDSQFSSEGSSECTERPPCTT 181

QY 347 KDYFQIHTPCDEEGKTQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPFGYNNGSS 406  
Db 182 KDYFQIHTPCDEEGKTQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPFGYNNGSS 241

QY 407 SCHPCPPGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGW 466  
Db 242 SCHPCPPGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGW 301

QY 467 EVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMGTATGSELGRITFTFETLCSADCVLV 526  
Db 302 EVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMGTATGSELGRITFTFETLCSADCVLV 361

QY 527 FMVDINRKSTNVVESWGTKEKQAYTHIIFKNATFTFTWAFQR 569  
Db 362 FMVDINRKSTNVVESWGTKEKQAYTHIIFKNATFTFTWGIPR 404

RESULT 11  
AAB35328  
ID AAB35328 standard; Protein; 750 AA.  
XX  
AC AAB35328;  
XX  
DT 08-MAY-2001 (first entry)  
XX  
DE Human TR13 receptor protein SEQ ID NO: 2.  
XX  
KW Human; tumour necrosis factor receptor; TR13; TR14; infection;  
KW cancer; autoimmune disease; allergy; inflammatory disease;  
KW graft rejection; apoptosis; cardiovascular disease; aneurysm.  
XX  
OS Homo sapiens.  
XX  
PN WO200105834-A1.  
XX



PD 25-JAN-2001.  
XX 14-JUL-2000; 2000WO-US19343.  
PF 16-JUL-1999; 99US-0144087.  
XX 18-AUG-1999; 99US-0149450.  
PR 20-AUG-1999; 99US-0149712.  
PR 10-SEP-1999; 99US-0153089.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA Ruben SM, Ni J, Young PE;  
XX WPI; 2001-112682/12.  
XX N-PSDB; AAF27997.  
DR Nucleic acids encoding 2 human tumor necrosis factor receptor  
XX polypeptides (TR13) and (TR14)), useful for the prevention, diagnosis  
PT and treatment of, e.g. cancers, acquired immune deficiency syndrome and  
PT hypohidrotic ectodermal dysplasia .  
XX Claim 40; Page 369-372; 418pp; English.  
PS The present invention provides the protein and coding sequences of the  
XX human tumour necrosis factor receptors TR13 and TR14. These sequences are  
CC useful in the diagnosis and treatment of many diseases, including cancer,  
CC autoimmune diseases, cardiovascular disorders, allergies,  
CC neurodegenerative diseases, graft rejection, inflammation, aneurysms and  
CC infections.  
XX Sequence 750 AA;  
SQ Query Match 37.3%; Score 2000; DB 22; Length 750;  
Best Local Similarity 53.2%; Pred. No. 2.5e-149;  
Matches 351; Conservative 111; Mismatches 186; Indels 12; Gaps 6;  
QY 308 QVCPRNTYSEKGAKEIRCKDDSQFS--GSSECTERPPCTTKDYFQIHTPCDEEGKTQIM 365  
Db 42 QTLPSNSYSNKGTSCHQC-DPDKYSEKSGSSCNVRPACTDKDYFHTACDANGETQLM 100  
QY 366 YKWIEPKICREDLTDAILRPPSGEKKDCPPCNPGFYNNGSSSCHPCPPGTFSDGTKCRP 425  
Db 101 YKWAKPKICSEDEGAVKLPASGVKTHCPCNPGFFKTNNSTCQPCPYGSYNGS-DCTR 159  
QY 426 CPAGTEPALGFYKWNVLPGNMKTSFCNVGNSKCDGMNGWEVAGDHQSGAGGSDNDYL 485  
Db 160 CPAGTEPAVGFEYKWNWTLPTNMTETVLGSGINFYKGMTGWVAGDHIYTAAGASDNDFM 219  
QY 486 ILNLHIPGFKPPTS-MTGATGSELGRITFVFETLCSADCVLVYFEMVDINRKSTNVVESWGG 544  
Db 220 ILTLVPGFRPQSVMAADTENKEVARITFVFETLCSVNCLEYFEMVGVNSRTNTPVETWKG 279  
QY 545 TKEQAYTHIIFKNATFTFTWAFQRTNQGDNRFRINDMKIYSITATNAVGVASSCRA 604  
Db 280 SKGKQSYTYIIEBNTTTSFTWAFQRTTFHEASRKYTNDVAKIYSINVTNVMNGVASYCRP 339  
QY 605 CALGSEQSGSCVPCPPGHYIEKETNOCKECPDPTYLSIHQVYGKEACIPCGPGSKNNOD 664  
Db 340 CALEASDVGSSTSCPAGYIIDRSDGTSCHSCPNTILKAHQYGVQACVPCPGPGTKNNKI 399  
QY 665 HVCYSDCFYHEKENQILHYDFNSLVSSVGLMNGPSETSKGTGYFHFNFISLCGHEGKK 724  
Db 400 HSLCYNDCTFSRNTPTRTFNYNFESALANTVTLAGGPSETSKGLKYFHFHTLSLCGNQGRK 459  
QY 725 MALCTNNITDFTVKEIVAGSDDYTNLVGAFCVQSTIIPSESKGFRAALSSQSIIADTFI 784  
Db 460 MSVCTDNVTDLRIPE---GESGFSKSITAYVCQAVIIPPEVTGYKAGVSSQPSVLADRLI 516  
QY 785 GVTVETTLKNNIKEDMFPVPTSOIPDVHFFYKSTATTSCINGRSTAVKMRCPNPTKSGA 844  
Db 517 GVTMTDLGITSAPAEFLHLESLGIPDVIFFRSNDVTQSCSSGRSTTIRVRCSPQKTVP 576  
QY 845 GVISVPSKCPAGTCDCGCTFYFLWESAECPLCTEHDFHETEGACKRGFQETLYVWNEPKW 904

Db 577 GSKLLPGTCSDGTCDCGNEHFLWESAAACPLCSVADYHAIVSSCVAGIQKTYVWREPKL 636  
QY 905 CIKGISLPEKKLATCETVDFWLKVAGVGAFATVALLVALTCYFVKKNQKKK---KTILN 960  
Db 637 CSGGISLPEQRVTICKTIDFWLKVGISAGTCTAILLTVLTCYFVKKNQKLEKYKSKLVMN 696  
RESULT 12  
AA59972  
ID AA59972 standard; Protein; 495 AA.  
XX AA59972;  
XX 31-JAN-2000 (first entry)  
XX Human endometrium tumour EST encoded protein 32.  
DE Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;  
XX treatment; uterine; gene therapy; expressed sequence tag.  
KW Homo sapiens.  
OS DE19817948-A1.  
XX 21-OCT-1999.  
XX 17-APR-1998; 98DE-1017948.  
XX 17-APR-1998; 98DE-1017948.  
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.  
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
XX WPI; 1999-591957/51.  
DR N-PSDB; AA241991.  
XX New nucleic acid sequences expressed in uterine cancer tissues, and  
PT derived polypeptides, for treatment of uterine and endometrial cancer  
PT and identification of therapeutic agents .  
XX Claim 23; Page 287; 444pp; German.  
XX This invention describes novel human nucleic acid (cDNA) sequences (A),  
CC that are highly expressed in uterine tumour tissue and which have  
CC anticancer and cytostatic activity. (A) are used (i) for recombinant  
CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
CC are used (i) to identify agents suitable for treatment of uterine or  
CC endometrial cancer; (ii) directly for treating these forms of cancer  
CC (including expression from gene therapy vectors) and (iii) for  
CC generation of specific antibodies. (A) are identified by assembling ESTs  
CC (expressed sequence tags) from a particular tissue type before comparison  
CC of expression patterns. This allows a significantly longer fragment of  
CC the gene to be revealed, so should reduce the number of failures  
CC associated with the fact that ESTs from different libraries may represent  
CC different parts of the same unknown gene, distorting the estimated  
CC frequency of occurrence in a particular tissue. AA59941-Y60328 represent  
CC protein fragments encoded by the human endometrium tumour cDNA library  
CC derived EST fragments represented in AA241981-Z42121.  
XX Sequence 495 AA;  
SQ Query Match 23.5%; Score 1258.5; DB 20; Length 495;  
Best Local Similarity 50.8%; Pred. No. 6.8e-91;  
Matches 219; Conservative 81; Mismatches 124; Indels 7; Gaps 2;  
QY 534 KSTNVVESWGGTKEQAYTHIIFKNATFTFTWAFQRTNQGDNRFRINDMKIYSITATN 593  
Db 2 RTNTPVETWKGSKGQSYTYIIEENTTTSFTWAFQRTTFHEASRKYTNDVAKIYSINVTN 61  
QY 594 AVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNOCKECPDPTYLSIHQVYGKEACI 653  
:::||||| || ||| : |||| || |||::: ||||| || ||| ||| : |||



Db 62 VMNGVASYCRPCALEASDVGSSTSCPAGYYIDRDSGTCHSCPPNTILKAHPYGVQACV 121

Qy 654 PCGPGSKNQDHSVCYDCFFYHEKENQILHYDFSNLSSVSGSLMNGPSTFTSKTKYFHHF 713

Db 122 PCGPGTKNKKIHSLCYNDCTFSRNTPTRTFNYNFESALANTVTLAGGPSFTSKGLKYFHHF 181

Qy 714 NISLCGHEGKKMALCTNITDFTVKEIVAGSDDYTNLVGAFVCQSTIIPSESKGFRRAALS 773

Db 182 TSLCGNQGRKMSVCTDNVTDLRIPE---GESGFSKSIYAVCQAVIIPPEVTGYKAGVS 238

Qy 774 SQSIILADTFIGVTVETTLKNINIKEDMEFPVPTSQIPDVHFFYKSSATTSCINGRSTAV 833

Db 239 SQPVSLADRLIGVTTDMTLDGITSPAELFHLESLGIPDVIFFFYRSNDVTQSCSSGRSTTI 298

Qy 834 KMRNCPKSGAGVISVPKCPAGTCDGCTFYFLWESAECPLCTEHDFHEIEGACKRGFQ 893

Db 299 RVRCSPOKTVPGSLLLPGTCSGTCDCGNFHFLESAAACPLCSVADYHAIIVSSCVAGIQ 358

Qy 894 ETLYVNEPKWCIKGISLPEKKLATCETVDFWLKVGAGVGCAFTAVLLVALTCYFWKKNQK 953

Db 359 KTYVWREPKLCSGGISLPEQRTICKTIDFWLKVGISAGTCTAILLTVLTCYFWKKNQK 418

Qy 954 KK----KTILN 960

Db 419 LEYKYSKLVNM 429

RESULT 13

AAB83853

ID AAB83853 standard; Protein; 383 AA.

XX

AC AAB83853;

XX

DT 23-JUL-2001 (first entry)

XX

DE Amino acid sequence of a human protein expressed in tumour cells.

XX

KW Tumour cell; immunological disease; autoimmune disease; cancer;

KW infection.

OS Homo sapiens.

XX

PN WO200131003-A1.

XX

PD 03-MAY-2001.

XX

PF 30-OCT-2000; 2000WO-FR03032.

XX

PR 29-OCT-1999; 99FR-0013629.

XX

PA (FABR ) FABRE MEDICAMENT SA PIERRE.

XX

PI Delneste Y, Magistrelli G, Jeannin P, Bonnefoy J;

XX

DR WPI; 2001-328651/34.

DR N-PSDB; AAF89777.

XX

PT New nucleic acid, expressed in tumours and lymphoid tissue is useful for

PT identifying agents for treating tumours and autoimmune disease

XX

PS Claim 10; Page 74-75; 85pp; French.

XX

CC The present sequence represents a human protein expressed in tumour

CC cells. The polynucleotide is useful for screening cDNA/genomic DNA banks

CC and for cloning isolated DNA; identifying mutant forms of the gene that

CC encodes a human protein, where the mutations are associated with

CC abnormal gene expression, or promoters and regulators of the gene,

CC particularly for diagnosis; for recombinant expression of the derived

CC protein; as probes and primers for detection and amplification; and

CC as antisense therapeutics. The tumour expressed protein is useful for

CC raising specific antibodies and to screen agents that modulate its

CC activity, bind to it or interact with it. These agents are potentially

CC useful for treatment or prevention of diseases associated with abnormal

CC expression/activity of the protein, particularly immunological diseases

CC (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic

CC infections.

XX

SQ Sequence 383 AA;

Query Match 22.6%; Score 1208.5; DB 22; Length 383;

Best Local Similarity 58.3%; Pred. No. 4.3e-87;

Matches 215; Conservative 53; Mismatches 94; Indels 7; Gaps 6;

Qy 167 CNNSSWIPRGNVIESNRDDCTVSLIYAVHLKKSGYVFFEYQYVDNNIFFEFFFQNDQCQE 226

Db 19 CTSSKWVPRGDYIASNTDECTATLMYAVNLKQSGTVNFYYYPDSSIIFEFFVQNDQCQP 78

Qy 227 MDTTDDKWVKLTNGEWGSHSVMLKSGTNNILYWRRTGILMGSKAVKPVLVKNITIEGVAY 286

Db 79 -NADDSRWMTTEKG-WEEHVSVELNRGNVLYWRTTAFSVWTKVPKPVLVNRNIAITGVAY 136

Qy 287 TSECPCPKPGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTERPPC 344

Db 137 TSECPCPKPGTYADKQSSFCCLKCPANSYSNKGTSCHQC-DPDKYSEKSSSSCNVRPAC 195

Qy 345 TTKDYFQIHTPCDEEKTQIMYKWIIEPKICREDLTDAIRLPPSGEKKDPPCNPGFYNNG 404

Db 196 TDKDYFYTHTACDANGETQLMYKWAKPKICSEDLLEGAVKLPASGVKTHCPCNPGFYKTN 255

Qy 405 SSSCHPCPPGTFSDGTRKCRPCPAGTEPALGFYKWNVNLPGNMKTSCFNVGNKCDGMN 464

Db 256 NSTCQPCPYGPYSNGS-DCTRCPAGTEPAVGFEYKWNLTPTNMETTSLGINFYKGMT 314

Qy 465 GWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTS-MTGATGSELGRITFVFETLCSADC 523

Db 315 GWEVAGDHIYTAAGASDNDFMILTLPVPGFRPPQSVYMADTENKEVARITEVFETLCSVNC 374

Qy 524 VLYFMVDIN 532

Db 375 ELIFMVGVN 383

RESULT 14

AAB85768

ID AAB85768 standard; Protein; 372 AA.

XX

AC AAB85768;

XX

DT 29-OCT-2001 (first entry)

XX

DE Human seven-transmembrane protein 50288 sequence.

XX

KW seven-transmembrane protein; G-protein coupled receptor; GPCR; human;

KW 17724; 50288; 31945; antiinflammatory; antiulcer; cytostatic; virucide;

KW hepatotropic; immunosuppressive; gynecological; neuroprotective;

KW anti-HIV; immunostimulant; dermatological; antiatherosclerotic; cardiant;

KW antianemic; antiparkinsonian; nephrotropic; antithyroid; hemostatic;

KW cerebroprotective; osteopathic; analgesic; gene therapy; nootropic.

XX

OS Homo sapiens.

XX

PN WO200159117-A2.

XX

PD 16-AUG-2001.

XX

PF 12-FEB-2001; 2001WO-US04536.

XX

PR 11-FEB-2000; 2000US-0182061.

XX

PA (MILL-) MILLENNIUM PHARM INC.

XX

PI Glucksmann MA, Silos-Santiago I;

XX

DR WPI; 2001-514670/56.

DR N-PSDB; AAH76195, AAH76196.

XX

PT New seven-transmembrane protein/G-protein coupled receptor polypeptides  
PT and polynucleotides for diagnosing, treating seven-transmembrane  
PT protein/receptor-related disorders and to identify modulators of  
PT therapeutic use  
XX  
PS Claim 8; Page 139-141; 144pp; English.  
XX  
CC The invention provides isolated seven-transmembrane protein/G-protein  
CC coupled receptor polypeptides selected from 17724, 50288, 31945 proteins.  
CC The polypeptides can be expressed by standard recombinant methodology.  
CC Modulators of the polypeptides can be identified using a competition  
CC binding assay or an assay for receptor-mediated signal transduction. The  
CC polypeptides and polynucleotides are useful as reagents or targets in  
CC seven-transmembrane protein/receptor assays applicable to treatment and  
CC diagnosis of seven-transmembrane protein/receptor-mediated disorders  
CC (see AAH76191 for a detailed description of the various disorders that  
CC can be treated or diagnosed using the polypeptides). The polynucleotides  
CC are useful to detect mutations in genes and gene expression products such  
CC as mRNA, as antisense constructs to control gene expression and for  
CC chromosome identification. The present sequence represents the human  
CC seven transmembrane protein 50288 sequence.  
XX  
SQ Sequence 372 AA;

Query Match 16.5%; Score 883; DB 22; Length 372;  
Best Local Similarity 51.7%; Pred. NO. 2.4e-61;  
Matches 163; Conservative 54; Mismatches 84; Indels 14; Gaps 7;

QY 23 RGRSPWPSPAWICWALACQAAWAG---DLPSSSRPLPPCQEKDYHFEYTECDSSGSR 79  
Db 15 RGRTERIPR---LWRL-----LLWAGTAFVQTQGTGPELHACKSEYHVEYTACDSTGSR 67  
QY 80 WRVAIPNSAVDCSGLDPPVRGKECTFSCASGEYLEMKNQVCCKGEGYSLGSGIKFDEW 139  
Db 68 WRVAVPHTPGLCTSLDPVKGTECSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEW 127  
QY 140 DELPAGFSNIATFMDTVVGPDSRDPDGCNNSWIPRGNYIESNRDDCTVSLIYAVHLKKS 199  
Db 128 DELPHGFASLSANMELDDSAES-TGNCTSSKWVPRGDYIASNTDECTATLMYAVNLKQS 186  
QY 200 GYVFFEYQYVDDNFFEFFIIONDQCQEMDTTDDKWVKLTNGEWGSHSVMLKSGTNILYW 259  
Db 187 GTVNFYYYPDSIIEFFVQNDQCP-NADDSRWMTTEKG-WEFHSVELNRGNVLYW 244  
QY 260 RTTGILMGSAVKPVLVKNITIEGVAYTSECFPPCKPGTFESNKPGEFNCQVCPRNTYSEKG 319  
Db 245 RTTAFSVWTKVPKPVLRNIAITGVAYTSECFPPCKPGTYADKQGSFCKLCPANSYSNKG 304  
QY 320 AKECIRCKDDSQFSG 334  
Db 305 ETSCHQC-DPDKYSG 318

RESULT 15  
AAB83852  
ID AAB83852 standard; Protein; 209 AA.

XX AAB83852;

AC AAB83852;

XX 23-JUL-2001 (first entry)

DT Amino acid sequence of a human protein expressed in tumour cells.

DE Tumour cell; immunological disease; autoimmune disease; cancer;

XX infection.

OS Homo sapiens.

XX WO200131003-A1.

PN 03-MAY-2001.

XX 30-OCT-2000; 2000WO-FR03032.

XX 29-OCT-1999; 99FR-0013629.  
PR (FABR ) FABRE MEDICAMENT SA PIERRE.  
XX  
PA Delneste Y, Magistrelli G, Jeannin P, Bonnefoy J;  
XX  
PI WPI: 2001-328651/34.  
XX N-PSDB; AAF89776.

DR New nucleic acid, expressed in tumours and lymphoid tissue is useful for  
DR identifying agents for treating tumours and autoimmune disease  
XX  
PT Claim 10; Page 71-72; 85pp; French.  
PT  
PS The present sequence represents a human protein expressed in tumour  
XX cells. The polynucleotide is useful for screening CDNA/genomic DNA banks  
CC and for cloning isolated DNA; identifying mutant forms of the gene that  
CC encodes a human protein, where the mutations are associated with  
CC abnormal gene expression, or promoters and regulators of the gene,  
CC particularly for diagnosis; for recombinant expression of the derived  
CC protein; as probes and primers for detection and amplification; and  
CC as antisense therapeutics. The tumour expressed protein is useful for  
CC raising specific antibodies and to screen agents that modulate its  
CC activity, bind to it or interact with it. These agents are potentially  
CC useful for treatment or prevention of diseases associated with abnormal  
CC expression/activity of the protein, particularly immunological diseases  
CC (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic  
CC infections.  
XX  
SQ Sequence 209 AA;

Query Match 13.2%; Score 705.5; DB 22; Length 209;  
Best Local Similarity 56.2%; Pred. NO. 1.2e-47;  
Matches 126; Conservative 26; Mismatches 55; Indels 17; Gaps 3;

QY 310 CPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEGKTOIMYKWI 369  
Db 2 CDPDKYSEK-----GSSSCNVRPACTDKDYFYHTACDANGETQLMYKWA 46  
QY 370 EPKICREDLTDAIRLPPSGEKKDCPPCNPFGYNNNGSSSCHPCPPPGTFSDGTECRPCPAG 429  
Db 47 KPKICSEDLEGAVKLPASGVKTHCPCPNPGFFKTNNSTCQCPYPGYSNGS-DCTRCPCAG 105  
QY 430 TEPALGFHEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNL 489  
Db 106 TEPAVGFHEYKWNVLTPTNMTETVLSGINFEYKGMTGWEVAGDHIYTAAGASDNDFMILT 165  
QY 490 HIPGFKPPTS-MTGATGSELGRITFVFETLCSADCVLYFMVDIN 532  
Db 166 VVPGFRPPQSVMAADTENKEVARITFVFETLCSVNCELYFMVGVN 209

Search completed: May 12, 2003, 13:18:51  
Job time : 51.4563 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:07:53 ; Search time 20.8085 Seconds  
(without alignments)  
1361.665 Million cell updates/sec

Title: US-10-073-333A-2  
Perfect score: 5357  
Sequence: 1 MLFRARGPVRRGWRPAEA.....TCYFWKKKQKKKKTILNLFN 963

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query %	Length	DB ID	Description
1	164	3.1	3111	2	US-08-460-309-4	Sequence 4, Appli
2	164	3.1	3111	2	US-08-125-077-4	Sequence 4, Appli
3	162	3.0	1253	3	US-08-479-722B-4	Sequence 4, Appli
4	155.5	2.9	1251	5	PCT-US95-02251-3	Sequence 3, Appli
5	155.5	2.9	1252	1	US-08-199-780-3	Sequence 3, Appli
6	155.5	2.9	1252	2	US-08-316-650-3	Sequence 3, Appli
7	151	2.8	2594	4	US-08-718-388-7	Sequence 7, Appli
8	151	2.8	5405	4	US-08-718-388-9	Sequence 9, Appli
9	149	2.8	610	1	US-08-365-470-3	Sequence 3, Appli
10	149	2.8	610	3	US-09-209-668-19	Sequence 19, Appli
11	149	2.8	610	4	US-09-009-490A-89	Sequence 89, Appli
12	148	2.8	2703	1	US-08-185-432-19	Sequence 19, Appli
13	148	2.8	2703	4	US-08-899-232-4	Sequence 4, Appli
14	145	2.7	610	6	5217870-2	Patent No. 5217870
15	145	2.7	1193	2	US-08-400-159-10	Sequence 10, Appli
16	145	2.7	1193	3	US-08-611-729A-10	Sequence 10, Appli
17	143.5	2.7	2523	1	US-08-185-432-18	Sequence 18, Appli
18	143.5	2.7	2523	4	US-08-899-232-3	Sequence 3, Appli
19	142.5	2.7	1940	2	US-08-644-271-30	Sequence 30, Appli
20	142.5	2.7	1940	4	US-09-077-955-34	Sequence 34, Appli
21	139	2.6	1219	4	US-08-882-046-5	Sequence 5, Appli
22	137.5	2.6	833	4	US-09-013-895A-5	Sequence 5, Appli
23	137.5	2.6	833	4	US-09-448-868-5	Sequence 5, Appli
24	137	2.6	1010	4	US-08-882-046-7	Sequence 7, Appli
25	137	2.6	1036	4	US-09-068-740A-6	Sequence 6, Appli
26	137	2.6	1187	4	US-09-068-740A-7	Sequence 7, Appli
27	137	2.6	1208	4	US-09-199-865-1	Sequence 1, Appli

28	137	2.6	1218	2	US-08-400-159-6	Sequence 6, Appli
29	137	2.6	1218	3	US-08-611-729A-6	Sequence 6, Appli
30	137	2.6	1218	4	US-08-882-046-2	Sequence 2, Appli
31	137	2.6	1218	4	US-09-214-278-7	Sequence 7, Appli
32	137	2.6	1218	4	US-09-068-740A-11	Sequence 11, Appli
33	137	2.6	1964	4	US-09-467-997-1	Sequence 1, Appli
34	136.5	2.5	2471	1	US-08-185-432-16	Sequence 16, Appli
35	136.5	2.5	2471	1	US-08-083-590A-19	Sequence 19, Appli
36	136.5	2.5	2471	3	US-08-532-384-19	Sequence 19, Appli
37	136.5	2.5	2471	4	US-08-899-232-1	Sequence 1, Appli
38	135.5	2.5	810	2	US-08-820-170A-34	Sequence 34, Appli
39	135.5	2.5	810	3	US-09-055-699-34	Sequence 34, Appli
40	135.5	2.5	810	4	US-09-273-565-34	Sequence 34, Appli
41	135.5	2.5	810	4	US-09-565-538-34	Sequence 34, Appli
42	135.5	2.5	810	4	US-09-661-468-34	Sequence 34, Appli
43	133.5	2.5	299	4	US-09-286-529-17	Sequence 17, Appli
44	133	2.5	300	2	US-08-794-796-2	Sequence 2, Appli
45	132.5	2.5	1198	4	US-09-245-041-131	Sequence 131, App

ALIGNMENTS

RESULT 1  
US-08-460-309-4  
; Sequence 4, Application US/08460309  
; Patent No. 5837496  
; GENERAL INFORMATION:  
; APPLICANT: Engvall, Eva  
; APPLICANT: Leivo, Ilmo  
; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin  
; TITLE OF INVENTION: Fragments and Uses Thereof  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/460,309  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/125,077  
; FILING DATE: 22-SEP-1993  
; APPLICATION NUMBER: US PCT/US 94/10730  
; FILING DATE: 21-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/472,319  
; FILING DATE: 30-JAN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/919,951  
; FILING DATE: 27-JUL-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 9721  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3111 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; US-08-460-309-4





Db 703 VNLESAVSPTDGSIAAAVEVCQPPGYTGSSCESCWPRHRRVNGTIFGGICEPCQCFGH 762  
QY 624 --YIEKETNOCKECPDPTYLSIHQVYGKEACIPCGPG-----SKNQDHSVCYSDCFFY 675  
Db 763 AESCDVGTGECLNCKDHT-----GGPYCDKCLPGFYGEPTKGTSEDCQPCACPLNIP 814  
QY 676 HEKENQILHYDFSNLSSVGLMNGPSFTSKGTYFHFFNISLC--GHEGKMAICT----- 729  
Db 815 SNNFSPTCHLD-----RSLGLICDG-----CPVGYTGPRCERCAEGYF 852  
QY 730 -----NNITDFTVKEIVAGSDDYTNLVGA-FVCQSTIIPSESKGFRAALS 773  
Db 853 QGPSVPGGSCQPCQCNLDIFS-----IPGSCD--SLGSCCLICKP-----GTTG 895  
QY 774 QSIIILADTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSATTSCINGRSTAV 833  
Db 896 RYCELCADGYFGDAVDA--KNCQ-----PCRCNAGGSFSEVCHSQTGQCEC---RANVQ 944  
QY 834 KMRCNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESAECPLCTEHDFHEIEGACKRGFQ 893  
Db 945 QORCD-----KCKAGT-----FGLQSARGCVPCNCNSFGSKSFDCEESGQ 984  
QY 894 ETLVWVNEPKWCIKGISLPEKKLATC 919  
Db 985 -----CWCQPGVT--GKKCDRC 999

RESULT 3

US-08-479-722B-4  
; Sequence 4, Application US/08479722B  
; Patent No. 6074840  
; GENERAL INFORMATION:  
; APPLICANT: Bonadio, Jeffrey  
; APPLICANT: Yin, Wushan  
; TITLE OF INVENTION: LATENT TGF( BINDING PROTEIN (LTBP)  
; TITLE OF INVENTION: GENES, COMPOSITIONS AND METHODS  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Williams, Morgan & Amerson  
; STREET: 7676 Hillmont, Suite 250  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77040  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/479,722B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US95/02251  
; FILING DATE: 21-FEB-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/316,650  
; FILING DATE: 30-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/199,780  
; FILING DATE: 18-FEB-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fussey, Shelley P.M.  
; REGISTRATION NUMBER: 39,458  
; REFERENCE/DOCKET NUMBER: 4100.000500/FUS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (713) 934-7000  
; TELEFAX: (713) 934-7011  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1253 amino acids  
; TYPE: amino acid

; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-479-722B-4  
Query Match 3.0%; Score 162; DB 3; Length 1253;  
Best Local Similarity 20.3%; Pred. No. 1.4e-05;  
Matches 150; Conservative 67; Mismatches 246; Indels .276; Gaps 48;  
QY 289 ECFPCPK--GTFSNKPFSFNCQVCP--NTYSEKGAKECIRCKDDSQFSGSSECTERPPC 344  
Db 601 EAEPGCGPGKGICMNTGGSYNCH-CNRGYRLHVAGGRSCV-----DLNECAKPHLC 650  
QY 345 TTKDY---FQIHTPCDEEGKTQIMY-----KWIEPKICREDLTDAIRLP---PSGEKKDC 393  
Db 651 GDGGFCINFPGHYKCN-----CYPGYRLKASRPPI-C-EDI-DECRDPSTCPDGKCNK 701  
QY 394 PPCNPGFYNNGSSSCHPCPPGTFSDGTKECR---PCPAGTEPALGFYKWNVLPGNMKT 450  
Db 702 P-----GSFKCIACQPGYRSQGGACRDVNECSGTPCSPG---WCENLPGSYRC 748  
QY 451 SC-----FNVGNSKC---DGMNGWEVAGDHIQSGAGSDNDYLILNLHIPGFKPPTSMTG 502  
Db 749 TCAQGIKTRTGRRLSCIDVDECEAGKVCQDGICTNTPGSFQCCLSGYHL----- 797  
QY 503 ATGSELGRITFV---FETLC-SADCVLYFMVDINRKSTNVVESW-----GGTK 546  
Db 798 --SRDRSRCEIDECDFPAACIGGDCI-----NTNGSYRCLCPLGHRVLVGRK 843  
QY 547 -----EKQAYTHIIFKNATFTFT---WAFQRTNQGDNRR-FIN- 581  
Db 844 CKKDIDECSDPGLCLPHACENLQGSYVCVDEGFTLTQDQHGCEEVEQPHHKKECYLNF 903  
QY 582 -DMVKIYSITATNAVDGVASSCRACALGSEQSGSSC--VPC-----PPGHYIE 626  
Db 904 DDTVFCDSVLATNVQ---QEC-CCSLGAGW-GDHCEIYPCPVYSSAEFHSVLPDGKRLH 958  
QY 627 KETNQCKECPDPTYLSIHQ--VYGKEACI--PCGPGSKNNQDHSVCYSDCFFYHEKENQI 682  
Db 959 SGQHQCELCIP-AHRDIDECILFGAEICKEGKC---VNTQPGYECYCKQGFY----- 1006  
QY 683 LHYDFSNLSSVGLMNGPSFTSKGTYFHFNFNISLCGHEGKKMALCTNNITDFTVKEIVA 742  
Db 1007 --YDGNLLECV-----DVDECLDESN---CRNGVCE----- 1032  
QY 743 GSDDYTNLVGAFVCQSTIIPSESKGFRAALSSQSIILADTFIGVTVETTLKNINIKEDMF 802  
Db 1033 -----NTRGGYRCACCT-PPAEYSPAQA---QCLI-----PERW 1061  
QY 803 PVPTSQIPDVHFFYKSSATTSCING-----RSTAVKMRGNP-TKSGAGVI 847  
Db 1062 STPQRDVKCA---GASEERTACVWGPWAGPALTFDDCCCRQPRLGTQCRPCPPRGTG-- 1115  
QY 848 SVPSKCPAGTCDGCTFYFLWESAECPLC-----TEHDFHEIEGACKRGFQETLYVWNE 901  
Db 1116 ---SQCPTSQSESNSF---WDTS---PLLLGKSPRDEDSSEDSDECR----- 1154  
QY 902 PKWCIKGISLPEKKLATCE 920  
Db 1155 ---CVSGRCVPRPGGAVCE 1170  
RESULT 4  
PCT-US95-02251-3  
; Sequence 3, Application PC/TUS9502251  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE  
; TITLE OF INVENTION: CELLS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston

```

; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02251
; FILING DATE: CONCURRENTLY HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/316,650
; FILING DATE: 30-SEP-1994
; CLASSIFICATION:
; APPLICATION NUMBER: US 08/199,780
; FILING DATE: 18-FEB-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UMIC009p--
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1251 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-02251-3

```

Query Match 2.9%; Score 155.5; DB 5; Length 1251;  
Best Local Similarity 20.2%; Pred. NO. 5.6e-05;  
Matches 149; Conservative 68; Mismatches 245; Indels 277; Gaps 48;

QY	289	ECFPCKP--GTFSNKPGSFNCQVCPR--NTYSEKGAKECIRCKDDSQFSGSSECTERPPC	344
Db	600	EAEPGCGKIGICMNTGGSYNCH-CNRGYRLHVGAGRSCV-----DLNECAKPHLC	649
QY	345	TTKDY---FQIHTPCDEEGKTQIMY-----KWIEPKICREDLTDAIRLP---PSGKKKDC	393
Db	650	GDGGFCINFPGHYKCN-----CYPGYRLKASRPIC-EDI-DECRDPSTCPDGKCNK	700
QY	394	PPCNPGFYNNGSSSCHPCPPPGTFSDGTKECR---PCPAGTEPALGFEYKWNVLPGNMKT	450
Db	701	P-----GSFKIACQPGYRSQGGGACRDVNECSEGTPCSPG---WCEKLPGSYRC	747
QY	451	SC-----FNVGNSKC---DGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTG	502
Db	748	TCAQGIRTRTGRLSCIDVDCEAGKVCQDGICITNTPGSFOCQCLSGYHL-----	796
QY	503	ATGSELGRITFV---FETLC-SADCVLYFMVDINRKSTNVVESW-----GGTK	546
Db	797	--SRDRSRCEIDECDFPAACIGGDCI-----NTNGSYRCLCPLGHLRVGGRK	842
QY	547	-----EKQAYTHIIFKNATFTFT---WAFQRTNQGDNR--FIN-	581
Db	843	CKKDIDECSDPGLCLPHACENLQGSYVCVDEGFTLTQDQHGCCVEEQPHHKKECYLNF	902
QY	582	DMVKIYSITATNAVDGVASSCRACALGSEQSGSSC--VPC-----PPGHVIE	626
Db	903	DDTVFCDSVLATNVNQ---QEC-CCSLGAGW-GDHCEIYPCPVYSSAEFHSVLPDGKRLH	957
QY	627	KETNQCKECPDLYLSIHQ--VYGKEACI--PCGPGSKNNQDHSVYCSDCFYHEKENQI	682
Db	958	SGQHQHCELCIP-AHRDIDECILFGAEICKEGKC----VNSQPGYECYCKQGFY-----	1005
QY	683	LHYDFSNLSSVSGSLMNGPSFTSKGTKYFHFFFNISLCGHGKKMALCTNNITDFTVKEIVA	742

```

Db      1006  --YDGNLLECV-----DVECLDES-----CRNGVCENTWR-----1035
QY      743  GSDDYTNLVGAFVQSTIIPSESKGFRAALSSQSIILADTFIGTVETT LKNINIKEDMF 802
Db      1036  -----LPCACT-PPAEYSPAQA---QCLI-----PERW 1059
QY      803  PVPTSQIPDVHFFYKSSATTSCING-----RSTAVKMRCNP-TKSGAGVI 847
Db      1060  STPQRDVKCA---GASEERTACVWGPWAGPALTFDDCCCRQRLGTQCRPCPRGTG-- 1113
QY      848  SVPSKCPAGTCDCGTFYFLWESAEACPLC-----TEHDFHEIEGACKRGFOETLYVWNE 901
Db      1114  ---SQCPTSQSESNSF---WDTs---PLLLGKSPREDSSEDSDECR----- 1152
QY      902  PKWCIKGISLPEKKLATCE 920
Db      1153  ---CVSGPCVPRPGGAVCE 1168

```

RESULT 5

```

US-08-199-780-3
; Sequence 3, Application US/08199780
; Patent No. 5763416
; GENERAL INFORMATION:
; APPLICANT: Bonadio, Jeffrey
; APPLICANT: Goldstein, Steven A.
; TITLE OF INVENTION: Gene Transfer Into Bone Cells
; TITLE OF INVENTION: And Tissues
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/199,780
; FILING DATE: 18-FEB-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UMIC:002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1252 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-199-780-3

```

Query Match	2.9%;	Score 155.5;	DB 1;	Length 1252;
Best Local Similarity	20.2%;	Pred. NO. 5.6e-05;		
Matches 149;	Conservative 68;	Mismatches 245;	Indels 277;	Gaps 48;

Qy	289	ECFPCKP--GTF\$NKP\$G\$FN\$CQVCPR--NTY\$EKGAKECIRCDD\$Q\$F\$G\$S\$S\$E\$C\$T\$E\$R\$P\$P\$C	344
Db	601	EAEP\$CG\$K\$G\$IC\$M\$T\$G\$S\$YN\$C\$-C\$N\$R\$Y\$R\$L\$H\$V\$G\$A\$G\$R\$S\$C\$V\$-----D\$L\$N\$E\$C\$A\$K\$P\$H\$L\$C	650
Qy	345	TTKDY--FQIHTPCDEEGKQIMY----KWIEPKICREDLTDAIRLP-----PSGEKKDC	393
Db	651	GDG\$F\$C\$IN\$F\$P\$G\$H\$Y\$K\$C\$N\$-----C\$Y\$P\$G\$Y\$R\$K\$A\$S\$R\$P\$P\$IC\$-E\$D\$I\$-D\$E\$C\$R\$D\$P\$S\$T\$C\$D\$G\$K\$C\$E\$N\$K	701



Db 1154 ---CVSGPCVPRPGGAVCE 1169

RESULT 7

US-08-718-388-7

; Sequence 7, Application US/08718388

; Patent No. 6271362

; GENERAL INFORMATION:

; APPLICANT: MORIKAWA, MINORU

; APPLICANT: HARADA, NAOKI

; TITLE OF INVENTION: GENE ENCODING IgG Fc REGION-BINDING

; TITLE OF INVENTION: PROTEIN

; NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH

; STREET: PO BOX 747

; CITY: FALLS CHURCH

; STATE: VA

; COUNTRY: USA

; ZIP: 22040-0747

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/718,388

; FILING DATE:

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: MURPHY JR, GERALD M

; REGISTRATION NUMBER: 28,977

; REFERENCE/DOCKET NUMBER: 0230-111

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 205-8000

; TELEFAX: (703) 205-8050

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2594 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-718-388-7

Query Match 2.8%; Score 151; DB 4; Length 2594;

Best Local Similarity 17.5%; Pred. No. 0.00047;

Matches 202; Conservative 103; Mismatches 346; Indels 502; Gaps 62;

QY 18 AEAPRRGRSPWPSPAWICCMALAGCQAAWAGDLPSS-----SRPLPPCQEKDYHFEYT 71

Db 765 AAAPNSCGRPCVEGCVC---LPGFVASGGACVPASSCGCTFQGLQLAPGQEV----- 814

QY 72 ECDSSGRWRVAIPNSAVDCSGLPDPV--RGKECTFCASGEVLEMKNQV-----119

Db 815 -----WADELQCRRCCTCNGATHQVTCRDQ---SCPAGERCSVQNGLLGCYPDRFGT 863

QY 120 CSKCGEGTYSLSGSIKGFDEWDELPAFGFSNIATFMDTVVGPSPDSRDPDGCNNSSWIPRGNVI 179

Db 864 CQSGSDPHYVSFDGRRFD-----FMGTCTYL--LVG-----SCGQNAALPAFRVL 906

QY 180 ESN--RDDCTVSLIYAVHLKSGY---VFFEYQ---YVDNNI--FFEFFIQNDCCQEMDTT 230

Db 907 VENEHRGSQTVSYTRAVRVEARGVKVAVRREYPGQVLVDVQLYLPEQAADGQVQVPROG 966

QY 231 TDKWVK---LTDNGEWGSH-SVMLKSGTNILYWRRTTGILMGSKAVKPVL-----VKNTT 280

Db 967 RDAVVRTDFGLTVTYDWNARVTAKEVPSSYAEALCGLCNGFNPGDPADDLALRGGGQAANAL 1026

QY 281 IEGVAYTSECFPCPKPGTFSNKPQSFNCQVCP-----RNT 314

Db 1027 AFGNSWQEE---TRPGCGATEPGD-----CPKLDLSVAQQLQSKNECGILADPKGPFREC 1078

QY 315 YSE---KGA-KECI--RCKDDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEGKTQIMYKW 368

Db 1079 HSKLDPQGAVRDCVYDRCLLPQG--SG-----PLCDALATYA--AACQAAGAT--VHPW 1126

QY 369 IEPKICREDLTDAIRLPPSGEKKDCP-----PCNPGFYNNNG 404

Db 1127 RSEELC-----PLSCPPPHSHVEACSYGCPLSCGDLVPVGGCGSECHGCVCDGFAISG 1180

QY 405 SSSCHP-----C-----PPG-TF-----SDGTKECRPCPAGTEPALGFEY 438

Db 1181 -ESCLPLASCGCVHQGTYPHPPGQTFYPGPGCDLSLCHCQEGGLVSCSSSCGPHEA----- 1234

QY 439 KWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPT 498

Db 1235 ----CQPSGGSGLGCVAVGSSTC-----QASGD-----PHY---T 1261

QY 499 SMTGATGSELGRITVFETLCSADCVLYFMVDINRKSTNVSVESWGTTKEKQAYTHII-FK 557

Db 1262 TFDGRRRDFMGTCVYVLAQTCGTRPGLHRAVLQE---NV--AWG--NGRVSVTRVITVQ 1314

QY 558 NATETF-----TW-----AFQRTNOGD-----NRRFINDMVKIY 587

Db 1315 VANFTLRLEQRQWKVTVNGVDMKLPVVLANGQIRASQHGSDVVIETDFGLRVAYDLV--Y 1372

QY 588 SITATNAVDGVASSCRACA-----LGSEQSG-----SSCV--- 617

Db 1373 YVRVTVPGNYQQMCGLCGNYNGDPKDDFKPNGSQAGNANEFNWSWEEVVPDSPCLPPT 1432

QY 618 PCPPGHYIEKETNQCKECPDPTVLSIHQVYGKEACIPCGPGSKNNQDHSVY----- 669

Db 1433 PCPPG---SEDCIPSHKCPPE-----LEKKYQKEEF--CGLLSSPTGPLSSCHKLVDPQGP 1483

QY 670 -SDCFEYHEKENQILHYDFSLSVSGSLMNGPSFTSKGTKYFHFNFNISLCGHEGKKMALC 728

Db 1484 LKDCIF-----DLCLGGGNLSILC 1502

QY 729 TNNITDFTVKEIVAGSDDYTNLVGAFF--CQST---IIPSESKGFRAALSSQSIILADTF 783

Db 1503 SN-----IHAYVSACQAGGHVEPWRTEF----- 1527

QY 784 IGVTVETTLKNINIKEDMFPVPTSIQIPDVHFFYKSSATTSCINGRSTAVKMRCNPTKSG 843

Db 1528 -----CPMECPPNSHYELCADTCSLGC----- 1549

QY 844 AGVISVPSKCPAGTCDG--CTEYFLWESAECPL----- 875

Db 1550 -SALSAPPQCQDGGCAEGCQCDSGLYNGQACVPIQQCGCYHNGVYYEPEQTVLIDNCROQ 1608

QY 876 -----CTEHD-----FHEI-----EGACKRG 891

Db 1609 CTCHAGKGMVCOEHSCKPGQVQCPSGGILSCVTKDPCGHGVTCTRPQETCKEQGGQGVCLPN 1668

QY 892 FQETLYVWNEPKW 904

Db 1669 YEATCWLWGDPHY 1681

RESULT 8

US-08-718-388-9

; Sequence 9, Application US/08718388

; Patent No. 6271362

; GENERAL INFORMATION:

; APPLICANT: MORIKAWA, MINORU

; APPLICANT: HARADA, NAOKI

; TITLE OF INVENTION: GENE ENCODING IgG Fc REGION-BINDING

; TITLE OF INVENTION: PROTEIN

; NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH

; STREET: PO BOX 747

; CITY: FALLS CHURCH

; STATE: VA

; COUNTRY: USA

; ZIP: 22040-0747



COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/718,388  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: MURPHY JR, GERALD M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 0230-111  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5405 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-718-388-9

Query Match 2.8%; Score 151; DB 4; Length 5405;  
Best Local Similarity 17.5%; Pred. No. 0.0015;  
Matches 202; Conservative 103; Mismatches 346; Indels 502; Gaps 62;

QY	18	AEAPRRGRSPWPSPAWICWALAGCAAWAGDLPSS-----SRPLPPCQEKDYHFEYT	71
Db	765	AAAPSNCSGRPCVEGVC---LPGFVASGGACVPASSCGCTFQGLQ LAPQEV-----	814
QY	72	ECDSSGRWRVAIPNSAVDCSGLPDVP--RGKECTFCASGEVLEMKNQV-----	119
Db	815	-----WADLCQRCTCNGATHQVTCRDQ---SCPAGERCSVQGLLCYPRFGT	863
QY	120	CSKCGESTYSLGSGIKFDEWDELPAFGSNIATFMDTVVGPSDSRDPDGCNNSSWIPRGNVI	179
Db	864	CQSGDPHYVSFDRREF-----FMGTCTYL--LVG-----SCGQNAALPAFRVL	906
QY	180	ESN--RDDCTVSLIYAVHLKSGY---VFHEYQ---YVDNNI--FFEFFIQNDQCQEMDTT	230
Db	907	VENHRGSQTVSYTRAVRVEARGVKVARRREYPGQVLVDVQLYLPFQAADGQVQVFRQG	966
QY	231	TDKWVK---LTDNGEWGSH-SYMLKSGTNILYWRTTGILMGSKAVKPVL-----VKNIT	280
Db	967	RDVVRTDFGLTVTYDWNARVTAKVPSYAEALCGLCGNFNGDPADDLALRGGGQAANAL	1026
QY	281	IEGVATSECFPCPKPGTFNSKPGSFNCQVCP-----RNT	314
Db	1027	AFGNSWQEE---TRPGCGATEPGD---CPKLDLSVAQLOSKNECGILADPKGPFREC	1078
QY	315	YSE---KGA-KECI--RCKDDSPFSGSSECTERPPCTTKDYFQIHTPCDEEGKTQIMYKW	368
Db	1079	HSKLDPQGAVRDCVYDRCLLPQG-SG-----PLCDALATYA--AACQAAGAT--VHPW	1126
QY	369	IEPKICREDLTDAILRPPSGEKKDCP-----PCNPGFYNNG	404
Db	1127	RSEELC-----PLSCPPHSHYACSYGCPPLSCGDLVPVPGCGSECHGCVDEGFALSG	1180
QY	405	SSSCHP-----C-----PPG-TF-----SDGFKECRPCPAGTEPALGFY	438
Db	1181	ESCLPLASCGCVHQGTYPHPPGQTFYPPGCDSLCHCQEGGLVSCSSSCGPHEA-----	1234
QY	439	KWNNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPT	498
Db	1235	----CQPSGGSLGCVAVGSSTC-----QASGD-----PHY---T	1261
QY	499	SMTGATGSELGRITFVFETLCSADCVLVFMVDINRKSTNVVESWGGTKEKQAYTHII-FK	557
Db	1262	TFDGRFRDFMGTCVYVLAQTCTGTPGLHRAVLQE---NV--AWG--NGRVSVTRVITVQ	1314
QY	558	NATFTF-----TW-----AFQRTNQGD-----NRRFINDMVKIY	587

Db	1315	VANFTLRLEQRQWKVTVNGVDMKLPVVVLANGQIRASQHGSDVVVETDFGLRVAYDLV--Y	1372
QY	588	SITATNAVDGVASSCRACA-----LGSEQSG-----SSCV---	617
Db	1373	YVRVTVPGNYYQOMCGLCGNYNGDPKDDFKPNGSQAGNANEFNGSWEEVVPDSPCLPPT	1432
QY	618	PCPPGHIETNOCKECPDPTYLSIHQVYGKEACIPCGPGSKNNQDHSVCY-----	669
Db	1433	PCPPG---SEDCIPSHKCPPE---LEKKYQKEEF--CGLLSPTGPLSSCHKLVDPPQGP	1483
QY	670	-SDCFFYHEKENQILHYDFSNLSSVSLMNGPSFTSKGTYHFFNFISLCGHEGKKMALC	728
Db	1484	LKDCIF-----DLCLGGGNLSILC	1502
QY	729	TNNITDFTVKEIVAGSDDYTNLVGAFV--CQST---IIPSESKGFRAALSSQSIILADTF	783
Db	1503	SN-----IHAYVSACQAAGGHVEPWRTETF-----	1527
QY	784	IGTVVETLKNINIKEDMFPVPTSQIPDVHFFYKSSSTATTSCTINGRSTAVKMRCNPTKSG	843
Db	1528	-----CPMECPNSHYELCADCATCSLGC-----	1549
QY	844	AGVISVPSKCPAGTCDG--CTFYFLWESAECPL-----	875
Db	1550	-SALSAPPQCQDGGCAEGCQCDSGLYNGQACVPIQCCGYHNGVYYPEQTVLIDNCRQQ	1608
QY	876	-----CTEHD-----FHEI-----EGACKRG	891
Db	1609	CTCHAGKGMVCOEHSCKPGQVCQPSGGILSCVTKDPCHGVTCRPOETCKEQGGQGVCLPN	1668
QY	892	FOETLYVWNEPKW 904	
Db	1669	YEATCWLWGDPHY 1681	

RESULT 9  
US-08-365-470-3  
; Sequence 3, Application US/08365470  
; Patent No. 5632991  
; GENERAL INFORMATION:  
; APPLICANT: Gimbrone, Jr., Michael A.  
; TITLE OF INVENTION: Antibodies Specific For E-selectin And The Uses  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX  
; STREET: 1100 New York Ave., NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/365,470  
; FILING DATE: herewith  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/102,510  
; FILING DATE: 05-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/850,802  
; FILING DATE: 13-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Markowicz, Karen R.  
; REGISTRATION NUMBER: 36,351  
; REFERENCE/DOCKET NUMBER: 0627.1350003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-2600

```
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-365-470-3

Query Match      2.8%; Score 149; DB 1; Length 610;
Best Local Similarity 19.6%; Pred. No. 7e-05;
Matches 136; Conservative 79; Mismatches 214; Indels 264; Gaps 40;

QY 29 WSPA--WI-----CCWALAGCAA-----WAGDLPSSSRPLPPC-----QEKDYH 67
Db 65 YSPSYWIGIRKVNNVVWVGTKPLTEEAKNWAPGEPNNRQKD-EDCVEIYIKREKDVG 123
QY 68 FEYTECDSSGSRWRVAIPNSA---VDCSGLPDPVVRGKECTFSCASGEYLEMKNQVCSKC 123
Db 124 MNW---DERCSKKLALCYTAACNTNTSCG-----HGEVETINNYTCKC 165
QY 124 GEGTSLGSGIKFDEWDELPAQFSNIATFMDTVVGPDSRDPG---CNNSSWIIPRGNYIE 180
Db 166 DPGF----SGLKCEQIVNCTA-----LESPEHGSVLCSH---PLGNF-- 200
QY 181 SNRDDCTVSLIYAVHLKKSGYVFEFYQYVDNNIEFFEQNDQCQEMDTTDDKWKVLTDN 240
Db 201 SYNSSCSISC-----DRGYL-----PSSMET-----MOCMSS 227
QY 241 GENGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTS-----ECFPCK 294
Db 228 GEWSA-----PIPACNV-VECDAVTNPANGFEVCFQ-N 258
QY 295 PGTFs-NKPGSFNCQ-----VCPRNTYSEKGAKECIRCKDDSQ-FSGSSECT 339
Db 259 PGSPFWNTTCTFDCCEEGFELMGAQSLQCTSSGNWDNEKPTCAVTCRAVRQPQNGSVRCS 318
QY 340 ERP-----PC--TTKDYFQIHTPCDEEGKTQIMYKWIEP-KICREDLTDAIRLPPSG 388
Db 319 HSPAGEFTFKSSCNFTCEEGFMLOGPAQVECTTQ--GQWTOQIPVCEAFQCTALSNPERG 376
QY 389 EKKDCPPCNPGEFYNNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNM 448
Db 377 -YMNCLPSASGSFERYGSSCFEFCQGVFLKGSRLQCGPTG-----EW-----DNE 421
QY 449 KTSFCNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMGTGSEL 508
Db 422 KPTCEAV---RCDVH-----QPPKGLVRCASPI 448
QY 509 GRITFVFETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTTWAFO 568
Db 449 GE--FTYKSSCAFSCEEFG-----ELYGSTQ-----LE 474
QY 569 RTNOGQDNRRFIN-DMVKIYSITATNAVDGVASS-----CR-ACALGSEQSGSCVPC 619
Db 475 CTSQGWTEEVPSQVVKCSSLAVPGKINMSCSGEPVFGTVCKFACPEGWTLNGSAARTC 534
QY 620 -PPGHVIEKETNQCKECPDPTVLSIHQVYGKEA 651
Db 535 GATGHW----SGLLPTCEAPTESNIPLVAGLSA 563

RESULT 10
US-09-209-668-19
; Sequence 19, Application US/09209668A
; Patent No. 6114517
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P.
; APPLICANT: Xu, Xiaoxing S.
; TITLE OF INVENTION: METHODS OF MODULATING TUMOR NECROSIS FACTOR
; TITLE OF INVENTION: alpha-INDUCED EXPRESSION OF CELL ADHESION MOLECULES
; FILE REFERENCE: ISPH-0336
; CURRENT APPLICATION NUMBER: US/09/209,668A
; CURRENT FILING DATE: 1998-12-10
```

```
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-209-668-19

Query Match      2.8%; Score 149; DB 3; Length 610;
Best Local Similarity 19.6%; Pred. No. 7e-05;
Matches 136; Conservative 79; Mismatches 214; Indels 264; Gaps 40;

QY 29 WSPA--WI-----CCWALAGCAA-----WAGDLPSSSRPLPPC-----QEKDYH 67
Db 65 YSPSYWIGIRKVNNVVWVGTKPLTEEAKNWAPGEPNNRQKD-EDCVEIYIKREKDVG 123
QY 68 FEYTECDSSGSRWRVAIPNSA---VDCSGLPDPVVRGKECTFSCASGEYLEMKNQVCSKC 123
Db 124 MNW---DERCSKKLALCYTAACNTNTSCG-----HGEVETINNYTCKC 165
QY 124 GEGTSLGSGIKFDEWDELPAQFSNIATFMDTVVGPDSRDPG---CNNSSWIIPRGNYIE 180
Db 166 DPGF----SGLKCEQIVNCTA-----LESPEHGSVLCSH---PLGNF-- 200
QY 181 SNRDDCTVSLIYAVHLKKSGYVFEFYQYVDNNIEFFEQNDQCQEMDTTDDKWKVLTDN 240
Db 201 SYNSSCSISC-----DRGYL-----PSSMET-----MOCMSS 227
QY 241 GENGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTS-----ECFPCK 294
Db 228 GEWSA-----PIPACNV-VECDAVTNPANGFEVCFQ-N 258
QY 295 PGTFs-NKPGSFNCQ-----VCPRNTYSEKGAKECIRCKDDSQ-FSGSSECT 339
Db 259 PGSPFWNTTCTFDCCEEGFELMGAQSLQCTSSGNWDNEKPTCAVTCRAVRQPQNGSVRCS 318
QY 340 ERP-----PC--TTKDYFQIHTPCDEEGKTQIMYKWIEP-KICREDLTDAIRLPPSG 388
Db 319 HSPAGEFTFKSSCNFTCEEGFMLOGPAQVECTTQ--GQWTOQIPVCEAFQCTALSNPERG 376
QY 389 EKKDCPPCNPGEFYNNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNM 448
Db 377 -YMNCLPSASGSFERYGSSCFEFCQGVFLKGSRLQCGPTG-----EW-----DNE 421
QY 449 KTSFCNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMGTGSEL 508
Db 422 KPTCEAV---RCDVH-----QPPKGLVRCASPI 448
QY 509 GRITFVFETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTTWAFO 568
Db 449 GE--FTYKSSCAFSCEEFG-----ELYGSTQ-----LE 474
QY 569 RTNOGQDNRRFIN-DMVKIYSITATNAVDGVASS-----CR-ACALGSEQSGSCVPC 619
Db 475 CTSQGWTEEVPSQVVKCSSLAVPGKINMSCSGEPVFGTVCKFACPEGWTLNGSAARTC 534
QY 620 -PPGHVIEKETNQCKECPDPTVLSIHQVYGKEA 651
Db 535 GATGHW----SGLLPTCEAPTESNIPLVAGLSA 563

RESULT 11
US-09-009-490A-89
; Sequence 89, Application US/09009490A
; Patent No. 6300491
; GENERAL INFORMATION:
; APPLICANT: Bennett and Mirabelli
; TITLE OF INVENTION: Oligonucleotide Modulation
; TITLE OF INVENTION: of Cell Adhesion
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Office of Jane Massey Licata
; STREET: 66 East Main Street
```





Best Local Similarity 18.8%; Pred. No. 0.00094;  
Matches 198; Conservative 99; Mismatches 338; Indels 418; Gaps 60;

```
QY 73 CDSSGSRWRVAIP-NSAVDCSGLPDPVR-----GKEC-----TFSCASGEYLEM 115
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 904 CDEDIDECSSPCRNAGASCLNVPGSYRCLCTKGYEGRDCAINTDDCASFPQNG----- 958
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 KNOVCSKCGEGTYSL-----GSGIKFDEWD-----ELPAGFS 147
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 959 --RTCLD-GIGDYSCLCVDFGDKHCETDINECLSQPCQNGATCSQYVNSYTCCTCPLGFS 1015
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 148 NIATFMDTVVGPDSRDPDGCNNSSWIPRGNVIES-NRDDCTVSLIYAVHLKSGYVFFEY 206
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1016 GI-----NCQTNDEDCTESSCLNGGSCIDGINGYNCSC-----LAGYSGANC 1057
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 207 QYVDNNIFFEFFFIONDQCQEMDTTDDKWVKLTDNGEWSHVMKSGTNILYWRRTTGILM 266
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1058 QYKLNKCDSNPCLNGATCHE-----QNNEYTCH-----CPSG-----FT 1091
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 267 GSKAVKPVLVKNITIEGVAYTSEC--FPCKPG-TFSNKPQSFNCQVCPRNTYSEKGAKEC 323
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1092 GKQCSE-----YVDMCGQSPCENGATCSQMKHQFSCK-CSAGWTGKLCVDVQT 1137
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 324 IRCKDDSQFSGSSECTERPPC---TTKDYFQIHT-----PCDE 358
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1138 ISQDAADRKGLS---LRQLCNNGTCKDYGNSHVCSQGYAGSYCQKEIDECQSQPCQN 1194
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 359 EGKTQIMYKWIPIKICREDLTDAIRLPPSGEK-----KDCPPCNPGFYNNGSSSCH---- 409
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1195 GGTCDRLIGAYECQ-CROGF-----QGONCELNIDDCAP-NP--CQNG-GTCHDRVM 1241
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 410 ----PCPPPGTF---SDGTKECRP--CPAGTEPALGFEYKWNVLPGNMKTSCFN-VGNS 458
|||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1242 NFSCSCPPTGTMGIICEINKDDCKPGAC-----HNGSCIDRVGGF 1281
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 459 KC---DGMNGWEVAGDHIQ-----SGAGGSDNDYLILNLHI---PGFKPPTSMTGATGS 506
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1282 ECVQCPGFVGARCEGDINECLSNPCSNAGTLDVCQLVNNYHCNCRPG----- 1328
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 507 ELGR-----ITFVFETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATF 561
: || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1329 HMGRHCEHKVDFCAQSPC-----QNGGNCNIRQSGHHICICNN--- 1365
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 562 TFTAFAFORTN---QGQDNRRFINDMVKIYSITATNAVDGVASSCRA--CALGSEQSGSSC 616
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1366 ----GFYGNCELSGQD-----CD--SNPCRVGNCVVADEGFGYRC 1400
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 617 VPCPPGHIYIEK-ETNQCKECPDPTY---LSIHQVYGKEACIPC----- 655
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1401 -ECPRGTLGEHCEIDTLDCESPNCAQGAACEDLLGDYECL-CPSKWKGRCDIYDANYP 1458
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 656 ----GPQSKNN-----QDHSVCYSDCFFYHEKENQILHYDFSNL 690
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1459 GWNGGSGSNDRYAADLEQQRAMCDKRGCTEKQKQNGICDSDCNTY-----ACNFDGND 1512
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 691 SSVGSLMNGPSTSGTKYFHEF-----NISLCGHEG-----KKMALCTNNITDFTV 737
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1513 ----SLGINPWANCTANECWNKFNKGCNEECNNAACHYDGHDCERKLKSCDTLFDAYCQ 1568
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 738 KEIVAGSDDYTNLVGAFCVQSTIIPSESKGFRAALSSQSTILADTFIGTVVETTLK---- 793
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1569 KHYDGFCDYG--CNNAECSWDGLDCENK-----TQSPVLAEGAMSVVMLMNVAPRE 1619
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 794 -----NINIKEDMF-----PVPTSQIPDVHFFYKKSSTATTSING 828
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1620 IQAQFLRNMSHMLRTTVRLKKDALGHDIIINWKDNVRVPEIETDFAFNKILYTOOVHQ 1679
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 829 RSTAVMRCNPTKSGAGVISVPSKCPAGTCDGCTFY-----FLWESAEACPLCTEHDHF 882
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1680 TGIQIYLEIDNRK-----CTECFTHAVEAAEFLAATAAKHQLRNDFOIH 1723
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 883 EIEGACKRGFQETLYVWNEP-----KWCIKGISL 911
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

Db 1724 SVRGIKNPGDEDN---GEPPANVKYVITGIIL 1752

RESULT 13  
US-08-899-232-4  
; Sequence 4, Application US/08899232  
; Patent No. 6436650  
; GENERAL INFORMATION:  
; APPLICANT: Artavanis-Tsakonas, Spyridon  
; APPLICANT: Qi, Huilin  
; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON  
; FILE REFERENCE: 7326-046  
; CURRENT APPLICATION NUMBER: US/08/899,232  
; CURRENT FILING DATE: 1997-07-23  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 2703  
; TYPE: PRT  
; ORGANISM: Drosophila sp.  
US-08-899-232-4

Query Match 2.8%; Score 148; DB 4; Length 2703;  
Best Local Similarity 18.8%; Pred. No. 0.00094;  
Matches 198; Conservative 99; Mismatches 338; Indels 418; Gaps 60;

```
QY 73 CDSSGSRWRVAIP-NSAVDCSGLPDPVR-----GKEC-----TFSCASGEYLEM 115
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 904 CDEDIDECSSPCRNAGASCLNVPGSYRCLCTKGYEGRDCAINTDDCASFPQNG----- 958
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 KNOVCSKCGEGTYSL---GSGIKFDEWD-----ELPAGFS 147
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 959 --RTCLD-GIGDYSCLCVDFGDKHCETDINECLSQPCQNGATCSQYVNSYTCCTCPLGFS 1015
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 148 NIATFMDTVVGPDSRDPDGCNNSSWIPRGNVIES-NRDDCTVSLIYAVHLKSGYVFFEY 206
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1016 GI-----NCQTNDEDCTESSCLNGGSCIDGINGYNCSC-----LAGYSGANC 1057
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 207 QYVDNNIFFEFFFIONDQCQEMDTTDDKWVKLTDNGEWSHVMKSGTNILYWRRTTGILM 266
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1058 QYKLNKCDSNPCLNGATCHE-----QNNEYTCH---CPSG-----FT 1091
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 267 GSKAVKPVLVKNITIEGVAYTSEC--FPCKPG-TFSNKPQSFNCQVCPRNTYSEKGAKEC 323
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1092 GKQCSE-----YVDMCGQSPCENGATCSQMKHQFSCK-CSAGWTGKLCVDVQT 1137
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 324 IRCKDDSQFSGSSECTERPPC---TTKDYFQIHT-----PCDE 358
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1138 ISQDAADRKGLS---LRQLCNNGTCKDYGNSHVCSQGYAGSYCQKEIDECQSQPCQN 1194
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 359 EGKTQIMYKWIPIKICREDLTDAIRLPPSGEK-----KDCPPCNPGFYNNGSSSCH---- 409
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1195 GGTCDRLIGAYECQ-CROGF-----QGONCELNIDDCAP-NP--CQNG-GTCHDRVM 1241
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 410 ----PCPPPGTF---SDGTKECRP--CPAGTEPALGFEYKWNVLPGNMKTSCFN-VGNS 458
|||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1242 NFSCSCPPTGTMGIICEINKDDCKPGAC-----HNGSCIDRVGGF 1281
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 459 KC---DGMNGWEVAGDHIQ-----SGAGGSDNDYLILNLHI---PGFKPPTSMTGATGS 506
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1282 ECVQCPGFVGARCEGDINECLSNPCSNAGTLDVCQLVNNYHCNCRPG----- 1328
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 507 ELGR-----ITFVFETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATF 561
: || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1329 HMGRHCEHKVDFCAQSPC-----QNGGNCNIRQSGHHICICNN--- 1365
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 562 TFTAFAFORTN---QGQDNRRFINDMVKIYSITATNAVDGVASSCRA--CALGSEQSGSSC 616
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1366 ----GFYGNCELSGQD-----CD--SNPCRVGNCVVADEGFGYRC 1400
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 617 VPCPPGHIYIEK-ETNQCKECPDPTY---LSIHQVYGKEACIPC----- 655
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1401 -ECPRGTLGEHCEIDTLDCESPNCAQGAACEDLLGDYECL-CPSKWKGRCDIYDANYP 1458
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 656 ----GPQSKNN-----QDHSVCYSDCFFYHEKENQILHYDFSNL 690
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1459 GWNGGSGSNDRYAADLEQQRAMCDKRGCTEKQKQNGICDSDCNTY-----ACNFDGND 1512
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 691 SSVGSLMNGPSTSGTKYFHEF-----NISLCGHEG-----KKMALCTNNITDFTV 737
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1513 ----SLGINPWANCTANECWNKFNKGCNEECNNAACHYDGHDCERKLKSCDTLFDAYCQ 1568
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 738 KEIVAGSDDYTNLVGAFCVQSTIIPSESKGFRAALSSQSTILADTFIGTVVETTLK---- 793
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1569 KHYDGFCDYG--CNNAECSWDGLDCENK-----TQSPVLAEGAMSVVMLMNVAPRE 1619
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 794 -----NINIKEDMF-----PVPTSQIPDVHFFYKKSSTATTSING 828
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1620 IQAQFLRNMSHMLRTTVRLKKDALGHDIIINWKDNVRVPEIETDFAFNKILYTOOVHQ 1679
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 829 RSTAVMRCNPTKSGAGVISVPSKCPAGTCDGCTFY-----FLWESAEACPLCTEHDHF 882
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1680 TGIQIYLEIDNRK-----CTECFTHAVEAAEFLAATAAKHQLRNDFOIH 1723
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 883 EIEGACKRGFQETLYVWNEP-----KWCIKGISL 911
: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```



QY	656	----	GP	G	S	K	N	N	-----	Q	D	H	S	V	C	Y	S	D	C	F	F	Y	H	E	K	E	N	Q	I	L	H	Y	D	F	S	N	L	690																					
						:	:	:			:	:					:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:																						
Db	1459	GW	G	G	S	G	S	G	N	D	R	Y	A	A	D	L	E	Q	R	A	M	C	D	K	R	G	C	T	E	K	Q	G	N	I	C	D	S	D	C	N	Y	-----	A	C	N	F	D	G	N	D	C	1512							
QY	691	SS	V	G	S	L	M	N	G	P	S	F	T	S	K	G	T	K	Y	F	H	F	-----	N	I	S	L	C	G	H	E	G	-----	K	K	M	A	L	C	T	N	N	I	T	D	F	T	V	737										
Db	1513	----	S	L	G	I	N	P	W	A	N	C	T	A	N	E	C	W	N	K	F	K	N	G	K	C	N	E	E	C	N	N	A	A	C	H	Y	D	G	H	D	C	E	R	K	L	K	S	C	D	T	L	F	D	A	Y	C	Q	1568
QY	738	KE	I	V	A	G	S	D	D	Y	T	N	L	V	A	F	V	C	Q	S	T	I	P	S	E	S	K	G	F	R	A	L	S	S	Q	S	I	I	L	A	D	T	F	I	G	V	T	V	E	T	L	K	-----	793					
Db	1569	K	H	Y	D	G	F	C	D	Y	G	----	C	N	N	A	E	C	S	W	D	G	L	D	C	E	N	K	-----	T	Q	S	P	V	L	A	E	G	A	M	S	V	V	M	L	M	N	V	E	A	F	R	E	1619					
QY	794	----	N	I	N	I	K	E	D	M	F	-----	P	V	P	T	S	Q	I	P	D	V	H	F	F	Y	K	S	T	A	T	T	S	C	I	N	G	828																					
Db	1620	I	A	Q	A	F	L	R	N	M	S	H	M	L	R	T	V	R	L	K	D	A	L	G	H	D	I	I	I	N	K	D	N	V	R	V	P	E	I	D	T	D	F	A	R	K	N	K	I	L	Y	T	Q	Q	V	H	Q	1679	
QY	829	R	S	T	A	V	K	M	R	C	N	P	T	K	S	G	A	G	V	I	S	V	P	S	K	P	A	G	T	C	D	G	C	T	E	F	Y	-----	F	L	W	E	S	A	E	A	C	P	L	C	T	E	H	D	F	H	882		
Db	1680	T	G	I	Q	I	Y	L	E	I	D	N	R	K	-----	C	T	E	C	F	H	A	V	E	A	A	E	F	L	A	A	T	A	A	K	H	Q	L	R	N	D	F	Q	I	H	1723													
QY	883	E	I	E	G	A	C	K	R	G	F	O	E	T	L	V	W	N	E	P	----	K	W	C	I	K	G	I	S	L	911																												
Db	1724	S	V	R	G	I	K	N	P	G	D	E	D	N	----	G	E	P	P	A	N	V	K	Y	I	T	G	I	L	1752																													

RESULT 14  
5217870-2  
; Patent No. 5217870  
; APPLICANT: HESSON, CATHERINE A.; LOBB, ROY R.; GOELZ, SUSAN E.  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES AGAINST CDX  
; NUMBER OF SEQUENCES: 4  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/345,151  
; FILING DATE: 28-APR-1989  
; SEQ ID NO: 2:  
; LENGTH: 610  
5217870-2

```

Db 377 -YMNCPLPSAGSFRYGSSCFESCQGFVLKGSKRQLCQGP TG-----EW-----DNE 421
QY 449 KTS CFNVGNSKCDGMNGWEVAGDHQSGAGGSDNDYLILNLHIPGFKPPTSM TGATGSEL 508
Db 422 KPTCEAV--RCDAVH-----PPPKGLVRCAHSPI 448
QY 509 GRITFEFETLCSADCVLVFMVDINRKSTNVVSWGGTKKEQAYTHIIFKNATFTFTWAFQ 568
Db 449 GE--FTYKSSCAFSCEE F-----ELHGSTQ-----LE 474
QY 569 RTNQGDNRFRIN-DMVKIYSITATNAVDGVASS-----CR-ACALGSEQSGSSCVPC 619
Db 475 CTSQGGWTEEVPSQVVKCSSLAVPGKINMSCSGEPVFGTVCKFACPEGWTLNGSAARTC 534
QY 620 -PPGHYIEKETNQCKECPDPDTYLSITHOVYGKEA 651
Db 535 GATGHW----SGLLPTCEAPTESNIP L VAGLSA 563

RESULT 15
US-08-400-159-10
; Sequence 10, Application US/08400159
; Patent No. 5869282
; GENERAL INFORMATION:
; APPLICANT: Ish-Horowicz, David
; APPLICANT: Henrique, Domingos M.P.
; APPLICANT: Lewis, Julian H.
; APPLICANT: Myat, Anna M.
; APPLICANT: Fleming, Robert J.
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Mann, Robert S.
; APPLICANT: Gray, Grace E.
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE
; TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/400,159
; FILING DATE: 07-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1193 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-400-159-10

```

QY 91 CSGLPDPVRGKECTFCASGEYLEMKNQVCCKGEGTYSLGSIGKFDENDELPAGFSNIA 15

Db 29 CDGTRNP-GDKKCT-----RDECDTYFKVCLKEYQSRVTAGG-----PCSFSGSKS 72

QY 151 TFMDTVVGPSDRPDGCNNSSWIPRGNYTESNRDDCTVSLIYAVHLKKSGYVF-FEYQYV 209

Db 73 T---PVIG-----GN-----TFNLKYSRNNKRNRIVIPFTFAWP 103

QY 210 DN-NIFFEFFIQDQCQEMDTTTKWKVLTLDNGEWGSHVMLKSGTNILYWRT-----TGI 264

Db 104 RSYTLLEAWDYNDNSTNPDRIEK-----ASHSGMINPSRQ---WQTLKHNTGA 150

QY 265 LMGSKAVKPVLVKNITIEGVAYTSECFP-CKPGTFSNKPGSFNCQVCPRNTYSEKGAKEC 323

Db 151 AHFEYQIR-----VTCAEHYYGFGCNKECRP-----RDDFFTHHTCDQN-----GNKTC 194

QY 324 IRCKDDSQFSGSSSECTE---RPPCTTKDYFIQIHTPCDEEGKTQIMYKW----- 368

Db 195 L-----EGWTG-PECNKAICRQGCSPK-----HGSCTVPGECRCQYGWQGQYCDKCIPHP 243

QY 369 -----IEP-----KICREDLTDAIRLPPSGEKKDCPPCNPGFYNNGSSSCHP 410

Db 244 GCVHGTCTIEPWQCLCETNWGGQLCDKDLNYCGTHPPCLNGGTCSNTGPDKY-----QC-S 297

QY 411 CPPGTFSDGTKECR-----PCPAG-----TEPALGFYKWNVLPGNMKTSCFNVGNSKCD 461

Db 298 CPEGYSGQNCETAEHACLSDPCHNGGSCLETSTGFE-----CV 335

QY 462 GMMGWEVAG-----DHIQS-----GAGGSDNDYLILNLHIPGFK--PPTSMTGATGSELG 509

Db 336 CAPGW--AGPTCTDNIDDCSPNPGCGHGTQCDL-----VDGFKCICPPQWTGKT----- 382

QY 510 RITVFETLCSAD-----CVLYFMVDINRKS-TNVVESWGGTKEKQAYTHIIFKNAT 560

Db 383 -----COLDANECEGKPCV-----NANSCRNLIGSY-----YDCI----- 413

QY 561 FTFTWAFQRTNOQDNRRFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSC----- 616

Db 414 --TGWS-----GHNC DININD-----CR-----GQCQNGGSCRDLY 442

QY 617 ----VPCPPGH---YIEKETNOCKECP-----PDYLSIHQVYVYKKEACIPCG-PGSKN 661

Db 443 NGYRCICSPGYAGDHCEKDINECASPNCMNGGHQCD-----EINGFQCLCPAGFSGNLC 496

QY 662 NQDHSVY-----SDCF-----FYHEKENQILHYDFSLSVG----- 694

Db 497 QLDIDYCEPNPCQNGAQCFLNAMYFCNCPED-----YEGKNCSHLKDHCRTPPCEVIDSC 552

QY 695 SLMNGPSFTSGTKYFHHFNISLCGHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAF 754

Db 553 TVAVASNSTPEGVRY---ISSNVCGPHGK---CKSQ-----AG-----GKF 587

QY 755 VCQSTIIPSESKGFRA-----ALSSQSIILADTFIGVTVETT 791

Db 588 TCEC-----NKGFTGTYPHENINDCESNPCKNGGTCTIDGVNSYKICSDGWEGTYCET- 640

QY 792 LKNINIKEDMPVPVPTSQIPDVHFFYKSSTATTSINGRSTAVKM-----RCNPTKSGAGV 846

Db 641 --NIN-----DCSKNPCHNGGTCTCRDLVNDFFCECKNGWKGKTC 676

QY 847 ISVPSKCPAGTC-----DGCTFYEL---WESAEACPL-----CTEHDFHE-----I 884

Db 677 HSRDSQCDEATCNNGGTCYDEGDTFKCMCPAGWEGA-TCNIARNSSCLPNPCHNGGTCTV 735

QY 885 EG-----ACKRGFQETLYVWN---EPKWCIGISLPEKKLATCETVDFWLKVAGVG 933

Db 736 SGDSFTCVCKEGWEGPTCTQNTNDCSHPHCYNS-----GTCVDGDNWYRCECAPG 785

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:21:38 ; Search time 176.631 Seconds  
(without alignments)  
501.729 Million cell updates/sec

Title: US-10-073-333A-2  
Perfect score: 5357  
Sequence: 1 MLFRARGPVGRGWRPAEA.....TCYFWKKNQKKKTILNLFN 963

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*

1:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2:	/cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7:	/cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10:	/cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12:	/cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5357	100.0	963	9	US-10-140-164-2
2	5307	99.1	1027	9	US-10-140-164-4
3	2796	52.2	1013	9	US-10-028-072-38
4	2796	52.2	1013	9	US-10-121-049-38
5	2796	52.2	1013	9	US-10-123-904-38
6	2796	52.2	1013	9	US-10-140-470-38
7	2796	52.2	1013	9	US-10-175-746-38
8	2796	52.2	1013	9	US-10-176-918-38
9	2796	52.2	1013	9	US-10-176-921-38
10	2796	52.2	1013	9	US-10-137-865-38
11	2796	52.2	1013	9	US-10-140-474-38
12	2796	52.2	1013	9	US-10-142-431-38
13	2796	52.2	1013	9	US-10-143-114-38
14	2796	52.2	1013	9	US-10-140-002-38
15	2796	52.2	1013	9	US-10-142-419-38
16	2796	52.2	1013	9	US-10-123-262-38
17	2796	52.2	1013	9	US-10-142-423-38
18	2796	52.2	1013	9	US-10-121-050-38
19	2796	52.2	1013	9	US-10-141-755-38

20	2796	52.2	1013	9	US-10-143-032-38	Sequence 38, Appl
21	2796	52.2	1013	9	US-10-123-108-38	Sequence 38, Appl
22	2796	52.2	1013	9	US-10-123-236-38	Sequence 38, Appl
23	2796	52.2	1013	9	US-10-123-261-38	Sequence 38, Appl
24	2796	52.2	1013	9	US-10-140-921-38	Sequence 38, Appl
25	2796	52.2	1013	9	US-10-140-928-38	Sequence 38, Appl
26	2796	52.2	1013	9	US-10-121-045-38	Sequence 38, Appl
27	2796	52.2	1013	9	US-10-123-292-38	Sequence 38, Appl
28	2796	52.2	1013	9	US-10-123-903-38	Sequence 38, Appl
29	2796	52.2	1013	9	US-10-124-819-38	Sequence 38, Appl
30	2796	52.2	1013	9	US-10-124-822-38	Sequence 38, Appl
31	2796	52.2	1013	9	US-10-140-925-38	Sequence 38, Appl
32	2796	52.2	1013	9	US-10-160-498-38	Sequence 38, Appl
33	2796	52.2	1013	9	US-10-121-041-38	Sequence 38, Appl
34	2796	52.2	1013	9	US-10-121-043-38	Sequence 38, Appl
35	2796	52.2	1013	9	US-10-121-047-38	Sequence 38, Appl
36	2796	52.2	1013	9	US-10-123-215-38	Sequence 38, Appl
37	2796	52.2	1013	9	US-10-123-902-38	Sequence 38, Appl
38	2796	52.2	1013	9	US-10-123-908-38	Sequence 38, Appl
39	2796	52.2	1013	9	US-10-123-909-38	Sequence 38, Appl
40	2796	52.2	1013	9	US-10-123-910-38	Sequence 38, Appl
41	2796	52.2	1013	9	US-10-124-813-38	Sequence 38, Appl
42	2796	52.2	1013	9	US-10-124-817-38	Sequence 38, Appl
43	2796	52.2	1013	9	US-10-124-824-38	Sequence 38, Appl
44	2796	52.2	1013	9	US-10-125-922-38	Sequence 38, Appl
45	2796	52.2	1013	9	US-10-125-924-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1  
US-10-140-164-2  
; Sequence 2, Application US/10140164  
; Publication No. US20030072736A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker et al.  
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16  
; FILE REFERENCE: PF514C1  
; CURRENT APPLICATION NUMBER: US/10/140,164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 963  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-140-164-2

Query Match 100.0%; Score 5357; DB 9; Length 963;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFRARGPVGRGWRPAEAPRRGRSPWPSPAWICCWALAGCQAAWAGDLPSSSSRPLPP 60

Db 1 MLFRARGPVGRGWRPAEAPRRGRSPWPSPAWICCWALAGCQAAWAGDLPSSSSRPLPP 60









```
RESULT 4
US-10-121-049-38
; Sequence 38, Application US/10121049
; Publication No. US2003002239A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C17
; CURRENT APPLICATION NUMBER: US/10/121,049
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 38
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 877, 882
; OTHER INFORMATION: unknown amino acid
US-10-121-049-38

Query Match      52.2%; Score 2796; DB 9; Length 1013;
Best Local Similarity 53.4%; Pred. No. 2.1e-205;
Matches 494; Conservative 159; Mismatches 254; Indels 18; Gaps 10;

QY 46 WAG---DLPSSSRPLPPCOEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKE 102
Db 31 WAGTAFQVQTGTGPELHACKSEYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKGTE 90

QY 103 CTFSCASGEYLEMKNQVCSKCGEGTYSLGSIGKIFDEWDELPAFGFSNIATFMDTVVGPDS 162
Db 91 CSFSCNAGEFLDMKDQCKPKCAEGRYSGLTGIRFDEWDELPHGFASLSANMELDDSAES 150

QY 163 RPDGCNNSWIPRGNVIESNRDDCTVSLIYAVHLKKSGYVFFFEYQYVDNIIFFEFFIQND 222
Db 151 -TGNCTSSKWVPRGDIYASNTDECTATLMYAVNLKQSGTVNFEYYPPDSSIIEFFVQND 209

QY 223 QCQEMDTTTDKWVKLLTDNGEWGSHSVMLKSGTNILYWRITGILMGSKAVKPVLVKNITIE 282
Db 210 QCQP-NADDSRWMTTEKG-WEFHSELNRGNVLYWRTTAFSVWTKVPKPVLVRNIAIT 267

QY 283 GVAYTSECFPCPKGTFSNKPQSFNCQVCPNRTYSEKGAKECIRCKDDSQFS--GSSECTE 340
Db 268 GVAYTSECFPCPKGTGYADKQGSFCKLCPANSYNKGETSCHQC-DPDKYSEKGSSCNV 326

QY 341 RPPCTTKDYFQIHTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPGE 400
Db 327 RPACTDKDYFYTHTACDANGETQLMYKWKAPKICSEDELEGAVKLPASGVKTHCPCNPGE 386

QY 401 YNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSFCFNVGNSKC 460
Db 387 FKTNNSTCQPCPYGSYNGS-DCTRCPAGTEPAVGFEYKWNWTLPTNMETVLVSGINFEY 445
```

```
QY 461 DGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTS-MTGATGSELGRITVFETLC 519
Db 446 KGMTGWEVAGDHIYTAAGASDNDNFILTLVVPGRPPQSVMDATENKEVARITVFETLC 505

QY 520 SADCVLVFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTTWFQRTNQGDNRFR 579
Db 506 SVNCELYFMVGVSRTNTPVETWKGSKGQSYTIIIEENTTSFTWAFQRTTFHEASRKY 565

QY 580 INDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPGPHYIEKETNOCKECPDPT 639
Db 566 TNDVAKIYSINVTNMVNGVASYCRPCALEASDVSSCTSCPAGYYIDRDSGTCHSCPPNT 625

QY 640 YLSIHQVYGKEACIPCGPGSKNNQDHSVYSDCFFYHEKENQILHYDFSNLSSVGLMNG 699
Db 626 ILKAHPYGVQACVPCPGPGTKNNKIHSLCYNDCTFSRNTPTRTFNYNFSALANTVTLAGG 685

QY 700 PSFTSKGTKYFHFNFNISLCGHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFCQST 759
Db 686 PSFTSKGLKYFHHFTLSLCGNQGRKMSVCTDNVTDLRIPE--GESGFSKSITAYVCQAV 742

QY 760 IIPSESKGFRAALSSQSIILADTFIGVTVETTLKNNIKEDMFPVPTSOIPDVHFFYKSS 819
Db 743 IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSAPLELFHLESLGIPDVIFFFYRSN 802

QY 820 TATTSICNGRSTAVKMRCPNPTKSGAGVISVPSKCPAGTCDGCTFFYFLWESAEACPLCTEH 879
Db 803 DVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCSDGTCDCGNFHFLEWESAAACPLCSVA 862

QY 880 DFHEIEGACKRGFQETLYVWNEPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAFTAVL 939
Db 863 DYHAIVSSCVAGIQXTTYVXREPKLCSGGISLPEQRVTICKTIDFWLKVGISAGTCTAIL 922

QY 940 LVALTCYFWKKNQKKK----KTILN 960
Db 923 LTVLTCYFWKKNQKLEYKYSKLVMN 947
```

```
RESULT 5
US-10-123-904-38
; Sequence 38, Application US/10123904
; Publication No. US20030022328A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C54
; CURRENT APPLICATION NUMBER: US/10/123,904
; CURRENT FILING DATE: 2002-04-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 38
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 877, 882
```



; OTHER INFORMATION: unknown amino acid									
US-10-123-904-38									
Query Match 52.2%; Score 2796; DB 9; Length 1013;									
Best Local Similarity 53.4%; Pred. No. 2.le-205;									
Matches 494; Conservative 159; Mismatches 254; Indels 18; Gaps 10;									
Qy	46	WAG--DLPSSSRPLPPCQEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLDPDPVRGKE	102						
Db	31	WAGTAFQVTQGTGPELHACKESYHYEYTACDSTGSRWRVAVPHTPGLCTSLSDPVKGTE	90						
Qy	103	CTFSCASGEYLEMKNQVCKGEGTYSLSGSIKFDEWDELPAGFSNIATFMDTVVGPSPDS	162						
Db	91	CSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAAES	150						
Qy	163	RPDGCNNSSWIPRGNYIESNRDDCTVSLIYAVHLKKSQYVFEFYQYVDNNIFFEFTQND	222						
Db	151	-TGNCCTSSKWWPRGDYIASNTDECTATLMYAVNLKQSGTVNFEYYPDSSIIFEFFVQND	209						
Qy	223	QCQEMDTTDDKWKVLTNDGEGSHSVMLKSGTNILYWRITGILMGSKAVKPVLVKNITIE	282						
Db	210	QCQP-NADDSRWMTTEKG-WEFHSVELNRGNVLYWRTTAFSVWTKVPKPVLRNIAIT	267						
Qy	283	GVAYTSECFPCPKPGTFESNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTE	340						
Db	268	GVAYTSECFPCPKPGTYADKQSSFCCLKCPANSYNKGETSCHQC-DPDKYSEKSSSCNV	326						
Qy	341	RPPCTTKDYFQIHTPCDEEGKTQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCPGF	400						
Db	327	RPACTDKDYFYTHTACDANGETQLMYKWKAPKICSEDLEGAVKLPASGVKTHCPCPNPGF	386						
Qy	401	YNNGSSSCHPCPPGTFSDGTEKCRPCPAGTEPALGFYKWNVLPGNMKTSCFNVNSKC	460						
Db	387	FKTNNSTCQPCPYGSYNGS-DCTRCPCPAGTEPAGVFYKWNVLPNTNMETTSLSGINFEY	445						
Qy	461	DGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTS-MTGATGSELGRITFEFTLC	519						
Db	446	KGMTGWEVAGDHIYTAAGASDNDFMILTLPVGPFRPPQSVMAADTENKEVARITFEFTLC	505						
Qy	520	SADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTWAFORTNQGDNRRF	579						
Db	506	SVNCELYFMVGVSRTNTPVETWKGSKGQSYTYIIEENTTSFTWAFORTTFHEASRKY	565						
Qy	580	INDMVKIYSITATNAVGVASSCRACALGSEQSGSSCVPCPPGPHYIEKETNQCKECPDPT	639						
Db	566	TNDVAKIYSINVTNVMNGVASYCRPCALEASDVSSCTSCPAGYIIDRSGTCHSCPNT	625						
Qy	640	YLSIHQVYKEACIPCGPGSKNNQDHSVCSYDCFFYHEKENQILHYDFSLSVSGSLMNG	699						
Db	626	ILKAHQYGVQACVPCPGPTKNNKIHSCLYNDCTFSRNTPTPTFNYSALANTVTLAGG	685						
Qy	700	PSFTSGTKYFHFENISLCGHEGKMACTNNITDFTVKEIVAGSDDYTNLVGAFVCQST	759						
Db	686	PSFTSKGLKYFHHFTLSLCGNQGRKMSVCTDNTVDLRIPE---GESGFSKSITAYVCOAV	742						
Qy	760	IIPSESKGFRAALSSQSIILADFTFGVTVETTLKNINIKEDMEPPVPTSQIPDVHFFYKSS	819						
Db	743	IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSAPAEHLHESLGIPDVIFFYRSN	802						
Qy	820	TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGTCDCGCTFYFLWESAECPLCTEH	879						
Db	803	DVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCSGDTCDGCFNHFLEWESAAACPLCSVA	862						
Qy	880	DFHEIEGACKRGFQETLYVWNEPKWCIKIGISLPEKKLATCETVDYFWLKVAGVGAFTAVL	939						
Db	863	DYHAIVSSCVAGIQXTTYVXREPKLCSGSLPEQRVTICKTIDFWLKVGISAGTCTAIL	922						
Qy	940	LVALTCYFWKKNQKKK---KTIILN 960							
Db	923	LTVLTCYFWKKNQKLEYKYSKLVMN 947							
RESULT 6									

US-10-140-470-38									
; Sequence 38, Application US/10140470									
; Publication No. US20030022331A1									
; GENERAL INFORMATION:									
; APPLICANT: Baker, Kevin P.									
; APPLICANT: Beresini, Maureen									
; APPLICANT: DeForge, Laura									
; APPLICANT: Desnoyers, Luc									
; APPLICANT: Filvaroff, Ellen									
; APPLICANT: Gao, Wei-Qiang									
; APPLICANT: Gerlitsen, Mary E.									
; APPLICANT: Goddard, Audrey									
; APPLICANT: Godowski, Paul J.									
; APPLICANT: Gurney, Austin L.									
; APPLICANT: Sherwood, Steven									
; APPLICANT: Smith, Victoria									
; APPLICANT: Stewart, Timothy A.									
; APPLICANT: Tumas, Daniel									
; APPLICANT: Watanabe, Colin K									
; APPLICANT: Wood, William									
; APPLICANT: Zhang, Zemin									
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC									
; TITLE OF INVENTION: ACIDS ENCODING THE SAME									
; FILE REFERENCE: P3330RLC160									
; CURRENT APPLICATION NUMBER: US/10/140,470									
; CURRENT FILING DATE: 2002-05-06									
; Prior Application removed - See Palm or File Wrapper									
; NUMBER OF SEQ ID NOS: 550									
; SEQ ID NO 38									
; LENGTH: 1013									
; TYPE: PRT									
; ORGANISM: Homo Sapien									
; FEATURE:									
; NAME/KEY: unsure									
; LOCATION: 877, 882									
; OTHER INFORMATION: unknown amino acid									
US-10-140-470-38									
Query Match 52.2%; Score 2796; DB 9; Length 1013;									
Best Local Similarity 53.4%; Pred. No. 2.le-205;									
Matches 494; Conservative 159; Mismatches 254; Indels 18; Gaps 10;									
Qy	46	WAG--DLPSSSRPLPPCQEKDYHFEYTECDSSSRWRVAIPNSAVDCSGLDPDPVRGKE	102						
Db	31	WAGTAFQVTQGTGPELHACKESYHYEYTACDSTGSRWRVAVPHTPGLCTSLSDPVKGTE	90						
Qy	103	CTFSCASGEYLEMKNQVCKGEGTYSLSGSIKFDEWDELPAGFSNIATFMDTVVGPSPDS	162						
Db	91	CSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAAES	150						
Qy	163	RPDGCNNSSWIPRGNYIESNRDDCTVSLIYAVHLKKSQYVFEFYQYVDNNIFFEFTQND	222						
Db	151	-TGNCCTSSKWWPRGDYIASNTDECTATLMYAVNLKQSGTVNFEYYPDSSIIFEFFVQND	209						
Qy	223	QCQEMDTTDDKWKVLTNDGEGSHSVMLKSGTNILYWRITGILMGSKAVKPVLVKNITIE	282						
Db	210	QCQP-NADDSRWMTTEKG-WEFHSVELNRGNVLYWRTTAFSVWTKVPKPVLRNIAIT	267						
Qy	283	GVAYTSECFPCPKPGTFESNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTE	340						
Db	268	GVAYTSECFPCPKPGTYADKQSSFCCLKCPANSYNKGETSCHQC-DPDKYSEKSSSCNV	326						
Qy	341	RPPCTTKDYFQIHTPCDEEGKTQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCPGF	400						
Db	327	RPACTDKDYFYTHTACDANGETQLMYKWKAPKICSEDLEGAVKLPASGVKTHCPCPNPGF	386						
Qy	401	YNNGSSSCHPCPPGTFSDGTEKCRPCPAGTEPALGFYKWNVLPGNMKTSCFNVNSKC	460						
Db	387	FKTNNSTCQPCPYGSYNGS-DCTRCPCPAGTEPAGVFYKWNVLPNTNMETTSLSGINFEY	445						
Qy	461	DGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTS-MTGATGSELGRITFEFTLC	519						
Db	446	KGMTGWEVAGDHIYTAAGASDNDFMILTLPVGPFRPPQSVMAADTENKEVARITFEFTLC	505						







QY 46 WAG---DLPSSSRPLPPCQEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKE 102  
Db 31 WAGTAFQVTOGTGPELHACKESEYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKGTE 90  
QY 103 CTFSCASGEYLEMKNQVCSKCGEGTYSLSGSIKFDWEDELPA GFSNIATFMDTVVGPSPDS 162  
Db 91 CSFSCNAGEFLDMKDQSCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAES 150  
QY 163 RPDGCNNSSWIPRGNYIESNRDDCTVSLIYAVHLKKSGYVFFEYQYVDNNIFFEFFIQND 222  
Db 151 -TGNCTSSKWVPRGDYIASNTDECTATLMYAVNLKQSGTVNFYIYPDSSIIEFFVQND 209  
QY 223 QCQEMDTTDDKWVKLTDNGEWGSHSVMLKSGTNILYWRTTGILMGSKAVKPVLVKNITIE 282  
Db 210 QCQP-NADDSRWKMTTEKG-WEFHVSVELNRGNVLYWRTTAFSVWTKVPKPVLRNIAIT 267  
QY 283 GVAYTSECFPCPKPGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTE 340  
Db 268 GVAYTSECFPCPKPGTYADKQGSSEFCCLCPANSYSNKGTSCHQC-DPDKYSEKSSSCNV 326  
QY 341 RPPCTTKDYFQIHTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSEKKKDCPPCNPGE 400  
Db 327 RPACTDKDYFYHTACDANGETQLMYKWKAPKICSEDLEGAVKLPA SGVKTCHCPCNPGE 386  
QY 401 YNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFEYKWWNVLPGNMKTSCFNVGNSKC 460  
Db 387 FKTNNSTCQCPYGSYNGS-DCTRCPAGTEPAVGFEYKWWNTLPTNMETTVLSGINFEY 445  
QY 461 DGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTS-MTGATGSELGRITFVFETLC 519  
Db 446 KGMTGWEVAGDHIYTAAGASDNDFMILT LVVPGFRPPQSVMA DTENKEVARITFVFETLC 505  
QY 520 SADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTTFAFQRTNQGDNRRE 579  
Db 506 SVNCELYFMVGNSRNTNTPVETWKGSKGQSYTYIIIEENTTTTSTFAFQRTTFHEASRKY 565  
QY 580 INDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNQCKECPDPT 639  
Db 566 TNDVAKIYSINVTNMVNGVASYCRPCALEASDVGSCTSCPAGYIIDRDSGTCHSCPNT 625  
QY 640 YLSIHQVYKGEACIPCGPGSKNNQDHSVCYSDCFYHEKENQILHYDFSNLSSVGLMNG 699  
Db 626 ILKAHQPIYGVQACVPCGPGTKNKNKIHSLCYNDCTFSRNTPTRTFNYSFALANTVTLAG 685  
QY 700 PSFTSKGTIFYHFFNISLCGHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFCQST 759  
Db 686 PSFTSKGLKYFHHFTLSLCGNQGRKMSVCTDNVTDLRIPE---GESGFSKSITAYVCQAV 742  
QY 760 IIPSESKGFRAALSSQSIIILADTFIGVTVETTLKNINIKEDMPVPVTSQIPDVHFFYKSS 819  
Db 743 IIPPEVTGKAGVSSQPVSLADRLIGVTTDMTLDGITS PAELFHLES LGIPDVIFFYRSN 802  
QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVP SKCPAGTCDGCTFYFLWESAECPLCTEH 879  
Db 803 DVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCS DGTCDGCPNPHFLWESAACPLCSVA 862  
QY 880 DFHEIEGACKRGFQETLYVWNEPKWCIKGISLPEKKLATCETVDFWLKVAGVGAFTAVL 939  
Db 863 DYHAIVSSCVAGIQXTTYVXREP KLSGGISLPEQRVTICKTIDFWLKVGISAGTCTAIL 922  
QY 940-LVALTCYFWKKNQKKK---KTI LN 960  
Db 923 LTVLTCYFWKKNQKLEYKYSKLVN 947

RESULT 10  
us-10-137-865-38  
; Sequence 38, Application US/10137865  
; Publication NO. US20030032155A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3330R1C154  
; CURRENT APPLICATION NUMBER: US/10/137,865  
; CURRENT FILING DATE: 2002-05-03  
; Prior Application removed - See Palm or File Wrapper  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 38  
; LENGTH: 1013  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 877, 882  
; OTHER INFORMATION: unknown amino acid  
us-10-137-865-38

Query Match 52.2%; Score 2796; DB 9; Length 1013;  
Best Local Similarity 53.4%; Pred. No. 2.1e-205;  
Matches 494; Conservative 159; Mismatches 254; Indels 18; Gaps 10;

QY 46 WAG---DLPSSSRPLPPCQEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKE 102  
Db 31 WAGTAFQVTOGTGPELHACKESEYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKGTE 90  
QY 103 CTFSCASGEYLEMKNQVCSKCGEGTYSLSGSIKFDWEDELPA GFSNIATFMDTVVGPSPDS 162  
Db 91 CSFSCNAGEFLDMKDQSCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAES 150  
QY 163 RPDGCNNSSWIPRGNYIESNRDDCTVSLIYAVHLKKSGYVFFEYQYVDNNIFFEFFIQND 222  
Db 151 -TGNCTSSKWVPRGDYIASNTDECTATLMYAVNLKQSGTVNFYIYPDSSIIEFFVQND 209  
QY 223 QCQEMDTTDDKWVKLTDNGEWGSHSVMLKSGTNILYWRTTGILMGSKAVKPVLVKNITIE 282  
Db 210 QCQP-NADDSRWKMTTEKG-WEFHVSVELNRGNVLYWRTTAFSVWTKVPKPVLRNIAIT 267  
QY 283 GVAYTSECFPCPKPGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTE 340  
Db 268 GVAYTSECFPCPKPGTYADKQGSSEFCCLCPANSYSNKGTSCHQC-DPDKYSEKSSSCNV 326  
QY 341 RPPCTTKDYFQIHTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSEKKKDCPPCNPGE 400  
Db 327 RPACTDKDYFYHTACDANGETQLMYKWKAPKICSEDLEGAVKLPA SGVKTCHCPCNPGE 386  
QY 401 YNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFEYKWWNVLPGNMKTSCFNVGNSKC 460  
Db 387 FKTNNSTCQCPYGSYNGS-DCTRCPAGTEPAVGFEYKWWNTLPTNMETTVLSGINFEY 445  
QY 461 DGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTS-MTGATGSELGRITFVFETLC 519  
Db 446 KGMTGWEVAGDHIYTAAGASDNDFMILT LVVPGFRPPQSVMA DTENKEVARITFVFETLC 505  
QY 520 SADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTTFAFQRTNQGDNRRE 579  
Db 506 SVNCELYFMVGNSRNTNTPVETWKGSKGQSYTYIIIEENTTTTSTFAFQRTTFHEASRKY 565  
QY 580 INDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNQCKECPDPT 639



Db 566 TNDVAKIYSINVTNVMNGVASYCRPCALEASDVGSSTSCPAGYYIDRDSGTCHSCPNT 625  
QY 640 YLSIHQVYKKEACIPCGSKNNQDHSVCYSDCFFYHEKENQILHYDFSNLSSVGLMNG 699  
Db 626 ILKAHQPYGVQACVPCGPGTKNNKIHSCLYNDCTFSRNTPTRTFNYNFSALANTVTLAGG 685  
QY 700 PSFTSKGTKYFHFNFISLCGHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFVCQST 759  
Db 686 PSFTSKGLKYFHFHTLSLQNGQGRKMSVCTDNVTDLRIPE---GESGFSKSITAYVQAV 742  
QY 760 IIPSESKGFRAALSSQSIILADTFIGVTETTLKNINIKEDMFPVPTSQIPDVHFFYKSS 819  
Db 743 IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSAPAELEHLESGLGIPDVIFYRSN 802  
QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEH 879  
Db 803 DVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCSDGTCDCGNFHFLESAAACPLCSVA 862  
QY 880 DFHEIEGACKRGFOETLYVWNEPKWCIKGISLPEKKLATCETVDVFWLKVAGVGAFVAVL 939  
Db 863 DYHAIVSSCVAGIQXTTYVXREPDKLCSGGLPEQRTVICKTIDFWLKVGISAGTCTAIL 922  
QY 940 LVALTCYFWKKNQKK---KTI LN 960  
Db 923 LTVLTCYFWKKNQKLEYKYSKLV MN 947

RESULT 11

US-10-140-474-38  
; Sequence 38, Application US/10140474  
; Publication No. US20030032156A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3330RLC162  
; CURRENT APPLICATION NUMBER: US/10/140,474  
; CURRENT FILING DATE: 2002-05-06  
; Prior Application removed - See Palm or File Wrapper  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 38  
; LENGTH: 1013  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 877, 882  
; OTHER INFORMATION: unknown amino acid  
US-10-140-474-38  
Query Match 52.2%; Score 2796; DB 9; Length 1013;  
Best Local Similarity 53.4%; Pred. No. 2.1e-205;  
Matches 494; Conservative 159; Mismatches 254; Indels 18; Gaps 10;

QY 46 WAG---DLPPSSSRPLPPCQEKDYHFEYTECDSSGSRWRVAIPNSAVDCSLPDPVRGKE 102  
||| : : : | :| :||| |||:|||||:| :| |||:|

Db 31 WAGTAFQVQTGTGPELHACKESEYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKGT E 90  
QY 103 CTFSCASGEYLEMKNQVCSKCGEGTYSLSGSIKFDWDELPA GFSNIATFMDTVVGP SDS 162  
Db 91 CSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIREDEWDELPHGFASLSANMELDDSAAES 150  
QY 163 RPDGCNNSWIPRGNVIESNRDDCTVSLIYAVHLKSGYVFF EYQYVDNNIFFE FFIQND 222  
Db 151 -TGNC TSSKWVPRGDYIASNTDECTATLMYAVNLKSGTVNF EYYPDSSIIFF E FVQND 209  
QY 223 QCQEMDTTDDKWVCLTDNGEWGSHSVMLKSGTNLIYWR TGTILMGSKAVKPVLVKNITIE 282  
Db 210 QCQP-NADDSRWMTTEKG-WEFHSELNRGNVLYWRTTAF SVWTKVPKPVLVRNIAIT 267  
QY 283 GVAYTSECFPCPKGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTE 340  
Db 268 GVAYTSECFPCPKGTYADKQSSFCCLKCPANSYSNKG ETSCHQC-DPDKYSEKGS S CNV 326  
QY 341 RPPCTTKDYFQIHTPCDEEKGTOIMYKWI EKPICREDLTD A IRLPPSGEKKDCPPCNP GF 400  
Db 327 RPACTDKDYFYTHTACDANGETQLMYKWKPKICSEDL EGAVKLPASGVKTHCPCPNP GF 386  
QY 401 YNNGSSSCHPCPPGTFSDGTPKECRPCPAGTEPALGFEYKWNVLPGNMKTS CFNVGNSKC 460  
Db 387 FKTNNSTCQPCPYGSYNGS-DCTRCPAGTEPAVGFEYKWN TLP TNMETTVLSGINPEY 445  
QY 461 DGMNWEVAGDHIQSGAGGSDNDYLILNLHI PGFKPPTS-MTGATGSELGRITFVFETLC 519  
Db 446 KGMTGEVAGDHIYTAAGASDNDNFILTLVVPGRPPQSVNADTENKEVARITFVFETLC 505  
QY 520 SADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTWAFORTNQGDNR RF 579  
Db 506 SVNCELYFMVGVSNTNTPTVETWKSGKGQSYTYII EENTTSFTWAFORTTTFHEASRY 565  
QY 580 INDMVKIYSITATNAV DGVASSCRACALGSEQSGSCVPCPPGHYIEKETNOCKECPD T 639  
Db 566 TNDVAKIYSINVTNVMNGVASYCRPCALEASDVGSSTSCPAGYYIDRDSGTCHSCPNT 625  
QY 640 YLSIHQVYKKEACIPCGPGSKNNQDHSVCYSDCFFYHEKENQILHYDFSNLSSVGLMNG 699  
Db 626 ILKAHQPYGVQACVPCGPGTKNNKIHSCLYNDCTESRNTPTRTFNYNFSALANTVTLAGG 685  
QY 700 PSFTSKGTKYFHFNFISLCGHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFVCQST 759  
Db 686 PSFTSKGLKYFHFHTLSLQNGQGRKMSVCTDNVTDLRIPE---GESGFSKSITAYVQAV 742  
QY 760 IIPSESKGFRAALSSQSIILADTFIGVTETTLKNINIKEDMFPVPTSQIPDVHFFYKSS 819  
Db 743 IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSAPAELEHLESGLGIPDVIFYRSN 802  
QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEH 879  
Db 803 DVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCSDGTCDCGNFHFLESAAACPLCSVA 862  
QY 880 DFHEIEGACKRGFOETLYVWNEPKWCIKGISLPEKKLATCETVDVFWLKVAGVGAFVAVL 939  
Db 863 DYHAIVSSCVAGIQXTTYVXREPDKLCSGGLPEQRTVICKTIDFWLKVGISAGTCTAIL 922  
QY 940 LVALTCYFWKKNQKK---KTI LN 960  
Db 923 LTVLTCYFWKKNQKLEYKYSKLV MN 947

RESULT 12  
US-10-142-431-38  
; Sequence 38, Application US/10142431  
; Publication No. US20030036179A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen





Db 91 CSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDSDAAES 150  
Qy 163 RPDGCNNSSWIPRGNYIESNRDDCTVSLIYAVHLKSKGYVFFEXQYVDNNIFFEFIQND 222  
Db 151 -TGNCCTSSKWVPRGDYIASNTDECTATLMYAVNLKQSGTVNFYYPPDSSIIFFEFVQND 209  
Qy 223 QCQEMDTTDDKWKVLTNDGEGSHSVMLKSGTNILYWRITGILMGSKAVKPVLVKNITIE 282  
Db 210 QCQP-NADDSRWMTTEKG-WEFHSVELNRGNVLYWRTTAFSVWTKVPKPVLVNRNIAIT 267  
Qy 283 GVAYTSECFPCPKPGTFSNKPQSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTE 340  
Db 268 GVAYTSECFPCPKPGTYADKQSSFCFKLCPANSYSNKGTSCHQC-DPDKYSEKGSSSCNV 326  
Qy 341 RPPCTTKDYFQIHTPCDEEGKTQIMYKWIIEPKICREDLTDALRPPSGEKKDCPPCPNGF 400  
Db 327 RPACTDKDYFYTHTACDANGETQLMYKWAKPKICSEDLEGAVKLPASGVKTHCPCPCNGF 386  
Qy 401 YNNGSSSCHPCPPGTFSDGTTKECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKC 460  
Db 387 FKTNNSTCQPCPYGSYNGS-DCTRPCAGTEPAVGFEYKWNWTLPTNMETTSLGINFEY 445  
Qy 461 DGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTS-MTGATGSELGRITFVFETLC 519  
Db 446 KGMTGWEVAGDHIYTAAGASDNDFMILTLPVPGFRPPQSVNADTENKEVARITFVFETLC 505  
Qy 520 SADCVLYFMVDINRKSTNVVSWGGTKEKQAYTHIIFKNATFTTFAFQRTNQGDNRRE 579  
Db 506 SVNCELYFMVGVSRTNTPVETWKSGKQSYTYIIIEENTTTSFTWAFQRTTFHEASRKY 565  
Qy 580 INDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGHYIEKETNQCKECPDPT 639  
Db 566 TNDVAKIYSINVTNMVNGVASYCRPCALEASDVSSCTSCPAGYIIDRDSGTCHSCPNT 625  
Qy 640 YLSIHQVYKKEACIPCGPGSKNQDHSVCYSDCFYHEKENQILHYDFSNLSSVGSMLMG 699  
Db 626 ILKAHQPYGVQACVPCPGPTKNNKTHSLCYNDCTFSRNTPTRTFNFSALANTVTLAG 685  
Qy 700 PSFTSKGTKYFHFNFNISLCGHEGKMACTNNITDFTVKEIVAGSDDYTNLVGAFVCQST 759  
Db 686 PSFTSKGLKYFHFHTLSLCGNQGRKMSVCTDNTVTLRIPE--GESGFSKSITAYVCOAV 742  
Qy 760 IIPSESKGFRAALSSQSIIADTFIGTVVETTLKNINIKEDMPVPTSQIPDVHFFYKSS 819  
Db 743 IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSPAELFHLESGLGIPDVIFFYRSN 802  
Qy 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGTCDCGCTFYFLWESAECPLCTEH 879  
Db 803 DVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPCTGSDGTCDGCNPFHFLWESAAACPLCSVA 862  
Qy 880 DFHEIEGACKRGFQETLYVWNEPKWCIKIGISLPEKKLATCETVDFWLKVGAGVGAFTAVL 939  
Db 863 DYHAIVSSCVAGIQXTTYVXREPKLCSGGISLPEQRVTICKTIDFWLKVGISAGTCTAIL 922  
Qy 940 LVALTCYFWKKNQKKK----KTIILN 960  
Db 923 LTVLTCYFWKKNQKLEYKYSKLVMN 947

RESULT 14

US-10-140-002-38

; Sequence 38, Application US/10140002  
; Publication NO. US20030037623A1

GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3330RIC59  
; CURRENT APPLICATION NUMBER: US/10/140,002  
; CURRENT FILING DATE: 2002-05-06  
; Prior Application removed - See Palm or File wrapper  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 38  
; LENGTH: 1013  
; TYPE: PRT  
; ORGANISM: Homo Saplen  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 877, 882  
; OTHER INFORMATION: unknown amino acid  
US-10-140-002-38

Query Match 52.2%; Score 2796; DB 9; Length 1013;

Best Local Similarity 53.4%; Pred. No. 2.1e-205;

Matches 494; Conservative 159; Mismatches 254; Indels 18; Gaps 10;

Qy 46 WAG---DLPSSSRRLPPCOEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKE 102  
Db 31 WAGTAFQVTQGTGPELHACKSEYHYEYTACDSTGSRWRVAVPHTPGLCTSLSDPVKGTGE 90  
Qy 103 CTFSCASGEYLEMKNQVCKGEGTYSLSGSIKGFDEWDELPAFESNIATFMDTVVGPDS 162  
Db 91 CSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDSDAAES 150  
Qy 163 RPDGCNNSSWIPRGNYIESNRDDCTVSLIYAVHLKSKGYVFFEXQYVDNNIFFEFIQND 222  
Db 151 -TGNCCTSSKWVPRGDYIASNTDECTATLMYAVNLKQSGTVNFYYPPDSSIIFFEFVQND 209  
Qy 223 QCQEMDTTDDKWKVLTNDGEGSHSVMLKSGTNILYWRITGILMGSKAVKPVLVKNITIE 282  
Db 210 QCQP-NADDSRWMTTEKG-WEFHSVELNRGNVLYWRTTAFSVWTKVPKPVLVNRNIAIT 267  
Qy 283 GVAYTSECFPCPKPGTFSNKPQSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTE 340  
Db 268 GVAYTSECFPCPKPGTYADKQSSFCFKLCPANSYSNKGTSCHQC-DPDKYSEKGSSSCNV 326  
Qy 341 RPPCTTKDYFQIHTPCDEEGKTQIMYKWIIEPKICREDLTDALRPPSGEKKDCPPCPNGF 400  
Db 327 RPACTDKDYFYTHTACDANGETQLMYKWAKPKICSEDLEGAVKLPASGVKTHCPCPCNGF 386  
Qy 401 YNNGSSSCHPCPPGTFSDGTTKECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKC 460  
Db 387 FKTNNSTCQPCPYGSYNGS-DCTRPCAGTEPAVGFEYKWNWTLPTNMETTSLGINFEY 445  
Qy 461 DGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTS-MTGATGSELGRITFVFETLC 519  
Db 446 KGMTGWEVAGDHIYTAAGASDNDFMILTLPVPGFRPPQSVNADTENKEVARITFVFETLC 505  
Qy 520 SADCVLYFMVDINRKSTNVVSWGGTKEKQAYTHIIFKNATFTTFAFQRTNQGDNRRE 579  
Db 506 SVNCELYFMVGVSRTNTPVETWKSGKQSYTYIIIEENTTTSFTWAFQRTTFHEASRKY 565  
Qy 580 INDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGHYIEKETNQCKECPDPT 639  
Db 566 TNDVAKIYSINVTNMVNGVASYCRPCALEASDVSSCTSCPAGYIIDRDSGTCHSCPNT 625  
Qy 640 YLSIHQVYKKEACIPCGPGSKNQDHSVCYSDCFYHEKENQILHYDFSNLSSVGSMLMG 699  
Db 626 ILKAHQPYGVQACVPCPGPTKNNKTHSLCYNDCTFSRNTPTRTFNFSALANTVTLAG 685







GenCore version 5.1.5  
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 11:39:43 ; Search time 28.0673 Seconds  
 (without alignments)  
 3298.405 Million cell updates/sec

Title: US-10-073-333A-2  
 Perfect score: 5357  
 Sequence: 1 MLFRARGPVRRGWRPAEA.....TCYFWKKNQKKKTILNLFN 963

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : PIR\_73:\*  
 1: pir1:\*  
 2: pir2:\*  
 3: pir3:\*  
 4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	242	4.5	1548	2 S34583	serine proteinase
2	210	3.9	713	2 A35502	major surface-labe
3	200	3.7	3635	2 T10053	laminin alpha 5 ch
4	198.5	3.7	677	2 C42125	trophozoite cystei
5	194	3.6	5376	2 T42215	zonadhesin - mouse
6	192.5	3.6	2823	2 T23064	hypothetical prote
7	192.5	3.6	2823	2 F87908	protein T22A3.8 [i
8	192.5	3.6	3102	2 T43291	laminin alpha chai
9	183	3.4	1372	2 T25933	hypothetical prote
10	182	3.4	667	2 A48579	trophozoite surfac
11	181	3.4	1766	2 A42125	trophozoite cystei
12	178.5	3.3	3672	2 T23433	hypothetical prote
13	178.5	3.3	3704	2 T37316	probable laminin a
14	177.5	3.3	1607	1 MMSB2	laminin gamma-1 ch
15	177.5	3.3	2907	2 A57278	fibrillin-2 precur
16	174.5	3.3	596	2 A45664	variant-specific s
17	173.5	3.2	2871	2 A55624	fibrillin-1 precur
18	173	3.2	1111	2 T26972	hypothetical prote
19	172	3.2	1786	1 MMSB1	laminin beta-1 cha
20	171.5	3.2	1620	2 T27283	hypothetical prote
21	170.5	3.2	2437	2 S42612	transmembrane prot
22	170	3.2	1557	2 T28811	hypothetical prote
23	168	3.1	1786	1 MMHUB1	hypothetical prote
24	166.5	3.1	557	2 A48434	laminin beta-1 cha
25	166	3.1	1274	2 T42017	variant-specific s
26	166	3.1	3002	2 A47221	cysteine rich prot
27	165	3.1	1639	1 MMFFB2	fibrillin 1 precur
28	164	3.1	2871	2 A55567	laminin gamma-1 ch
29	164	3.1	2918	2 A54105	fibrillin I - bovi
					fibrillin-2 precur

30	164	3.1	3084	1 MMSA	laminin alpha-1 ch
31	163	3.0	1297	2 T30274	proteolialisin - se
32	161.5	3.0	1790	1 MMFFB1	laminin beta-1 cha
33	161	3.0	3712	2 S18253	laminin alpha-1 ch
34	160.5	3.0	1609	1 MMHUB2	laminin gamma-1 ch
35	159.5	3.0	2352	2 T30201	Notch homolog prot
36	156.5	2.9	3106	1 S53868	laminin alpha-2 ch
37	153.5	2.9	1680	2 A43434	furin (EC 3.4.21.7
38	152.5	2.8	1251	2 A57293	latent transformin
39	152	2.8	1820	2 A55494	latent transformin
40	152	2.8	3707	2 S18252	heparan sulfate pr
41	151.5	2.8	2824	2 T22759	hypothetical prote
42	150.5	2.8	2703	1 A24420	notch protein - fr
43	149	2.8	1203	2 A49175	notch B protein -
44	149	2.8	1299	2 T43251	furin (EC 3.4.21.7
45	148.5	2.8	2491	1 A28372	insulin-like growt

ALIGNMENTS

RESULT 1  
 S34583

serine proteinase (EC 3.4.21.-) PC6B - mouse

C;Species: Mus musculus (house mouse)

C;Date: 02-Dec-1993 #sequence\_revision 10-Nov-1995 #text\_change 05-Nov-1999

C;Accession: S34583

R;Nakagawa, T.; Murakami, K.; Nakayama, K.

FEBS Lett. 327, 165-171, 1993

A;Title: Identification of an isoform with an extremely large Cys-rich region of PC6,  
 A;Reference number: S34583; MUID:93327934; PMID:8335106

A;Accession: S34583

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1548 <NAK>

A;Cross-references: GB:D17583; NID:g407344; PIDN:BAA04507.1; PID:d1005033; PID:g44037  
 C;Keywords: hydrolase; serine proteinase

Query Match 4.5%; Score 242; DB 2; Length 1548;  
 Best Local Similarity 19.1%; Pred. No. 8.2e-08;  
 Matches 221; Conservative 97; Mismatches 382; Indels 456; Gaps 64;

QY	61	COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKEC-----TFSCA 108
Db	347	CPPGHYHADKKRC-----RKCAPN-CESCFG----SHGQCLSKYGYFLNEETSSCV 394
QY	109	S-----GEYLEMKNQVCSKCGEGYSLGSGIKFDEWDELPAFGS-----NIATFMDTVVG 158
Db	395	TQCPDGSYEDIKKNVCGKCSENCKAC---IGFHNCTECKGGLSLQGSRCSVTCEDGQFFN 451
QY	159	PSDSRP-----DGCNNSSWIPRGNYESNR--DDCTVSLIYAVHLKKSGYVF 203
Db	452	GHDCQPCHRFCATCSGAGADGCINCT---EGYVMEEGRCVQCSVS-YVLDHSSEGGYK- 506
QY	204	FEYQYVDNNIFF-----EFFIQNDQCQEMDTT 231
Db	507	-SCKRCDNSCLTCNGPGFKNCSSCPSGYLLDLGTQCMGAICKDGEYIDDGHCQTCEASC 565
QY	232	DKWKILTDNGEWSHVMKSGTNTLYWRTTGILMGSKAVKPVLVKNITIEGVAYTSECF 291
Db	566	AKC-----WG-----PTQEDCISCPVTRVLDGRC-----VMNCPSWKFFFKKQCH 606
QY	292	PCKPGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDD----SQFSGSSECTERPPCTTK 347
Db	607	PC-----HYTCQGC-----QSGGPSNCTSCRADKHGQERFLYHGECLEN--CPVG 649
QY	348	DY-FQIHT--PCDEEGKTQIMYKWIPEKPKICREDLTDAIRLPSS-----GE 389
Db	650	HYPAGKHTCLPCPD--NCELCY---NPHICSRMSGYVIIPNHTCQKLECRQGEFQDSE 704
QY	390	KKDCPPCNPFGFY---NNGSSSCHPCPPGTFSDGTKECRPCPACTEPALGFYKWMNVLPG 446
Db	705	YEECMPCEEGLGCTEDDPGACTSCATGYIMFERHCYKACPEKT-----FGVKWECRACG 759



A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-3635 <MIN>  
A;Cross-references: EMBL:U37501; NID:g2599231; PID:g2599232  
C;Genetics:  
A;Gene: Lama5  
C;Superfamily: unassigned EGF-related proteins; EGF homology; laminin-type EGF-like homology  
C;Keywords: basement membrane; cell binding; extracellular matrix  
F;1888-1939/Domain: laminin-type EGF-like homology <LEG>  
F;1942-1970/Domain: EGF homology <EGF>

Query Match 3.7%; Score 200; DB 2; Length 3635;  
Best Local Similarity 18.5%; Pred. No. 0.00013;  
Matches 174; Conservative 82; Mismatches 279; Indels 404; Gaps 46;

QY	6	RGVVRGRG-----WGRPAEAPRRGRSPWPSPAWICCWALAGCQAAWAGDLPSS---- <td>54</td>	54
Db	1382	RGHVIGRDCSRCATGYWGFNCRP-----CDCGARLDELGTGCICP	1423
QY	55	SRPLPP---CQEKDYHFE---YTECDSSG-SRWRVAIPNSAVD---CSGLPDPVRGKE	102
Db	1424	PRTPVPDCLVCQPQSFQSGCHPLVGCCECNCSGPGVQELTDPCTDMDSGQCRCPN-VAGRR	1482
QY	103	CYFSCASGEY-----LEMKQV-----CSKCGEGTYSLGS	132
Db	1483	CD-TCAPGFYGYPSRCPCDCHCACTMASVCDPLTGTQCHCKENVQSGRCDQCVRVGTFSLDA	1541
QY	133	G-----IKFDEWDELPAFGSNIATF-----MDTVVGPDSDRP-----	164
Db	1542	ANPKGTRCFQGATER--CGNSNLARHEFVDMEGWVLLSSDRQVVPHEHRPEIELLHAD	1599
QY	165	-----DGCNNSWIPRGNIESNRDDCTVSLIYAVHLK-KSGYVFFEYQYVDNNIFFEFF	218
Db	1600	LRSVADTFSELYWQAPPSYLGDRVSSYGGTLHYELHSETQRGDIPIYESRPDVV-----	1654
QY	219	IQNDQCQ-----EMDTTDTKWVKL-TDNGEWSHVMKSGTNILYWR	260
Db	1655	LOGNQMSIAFLELAYPPPGQVHRGQLQLVEGNFRHLETHNPVSREELMMVLAGLEQLQIR	1714
QY	261	TTGILMGSKAVKPVLVKNITIEGVAYT-----SECFPCPKPGTFESN	300
Db	1715	A-----LFSQTSVVSLRRVVLEVASEAGRPPASNVELCMCPANYRGDSQCECAPGYRD	1770
QY	301	KPGSF--NCQVCPRNTYSEK---GAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHFP	355
Db	1771	TKGLFLGRVCVPCQCHGSHDRCLPGSGICVGCQHNTEGDAQERC--RPGFVSSDPSNPASP	1828
QY	356	CD-----EEGKTQIMYKWIPEKPKICREDLTDAIRLPPSGEKKDCPPC	396
Db	1829	CVSCPCPLAVPSNNFADGCVLRNGRTQC-----LCR-----PGYAGASCERC	1870
QY	397	NPGFYNNG---SSSCHPC-----PPGTFSDG---TKECRPC-----PAGTEPALGFE	437
Db	1871	APGFFGNPLVLGSSCQPCDCSGNGDPNMIFSDCDPLTGACRCLRHTTGPHCERCAPGF-	1929
QY	438	YKWWN-VLPGN-MKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGEK	495
Db	1930	--YGNALLPGNCTRCDCSPCGTETCDPQSG-----	1957
QY	496	PPTSMTGATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHII	555
Db	1958	-----RCLCKAGVT-----GQRCDRCLEGYFGFEQCQ-----	1984
QY	556	FKNATFTTWFQRTNQGDNRFRFINDVMVKIYSITATNAVDGVASSCRACALGSEQSGSS	615
Db	1985	-----GCRPCACGPAAKGSE	1999
QY	616	CVP-----CPPGHY--IEKETNQCK-----ECPP-----	637
Db	2000	CHPQSGQCHCQPGTTGPQCLECAPGYWGLPEKGCRRRCQCQPRGHCDPHTGHCTCPPGLSGE	2059
QY	638	--DTYLSIHQYVGKEACIPCPCGSKNNQ-----DHSVVCYSCFFYHEKENQILHYDFSMLS	691

Db	2060	RCDTCSQOHQV-----PVPKPGGGHGHCEVCDHCV-----VLLLD--DLE	2098
QY	692	SVGSL-----MNGPSFTSKGTYKYPHFENISLCGHEGK	723
Db	2099	RAGALLPAIREQLQGINASSAAWARLHRLNASIADLQSK	2137

RESULT 4  
C42125  
trophozoite cysteine-rich surface antigen 72 - Giardia lamblia (fragment)  
N;Alternate names: CRP72  
C;Species: Giardia lamblia  
C;Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 05-Dec-1998  
C;Accession: C42125  
R;Adam, R.D.; Yang, Y.M.; Nash, T.E.  
Mol. Cell. Biol. 12, 1194-1201, 1992  
A;Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170  
A;Reference number: A42125; MUID:92186850; PMID:1545800  
A;Accession: C42125  
A;Molecule type: DNA  
A;Residues: 1-677 <ADA>  
A;Cross-references: GB:M83934; NID:gl59123  
A;Experimental source: trophozoites  
A;Note: sequence extracted from NCBI backbone (NCBIN:88443, NCBIIP:88444); this ORF is  
C;Keywords: surface antigen

Query Match 3.7%; Score 198.5; DB 2; Length 677;  
Best Local Similarity 17.4%; Pred. No. 2.3e-05;  
Matches 181; Conservative 87; Mismatches 268; Indels 507; Gaps 47;

QY	7	GPVGRGWGRPAEAPRRGRSPWPSPAWICCWALAGCQAAWAGDLPSSSRPLPPCQE-KD	65
Db	35	GGVPVDGFCRPFQSPQA-----AAAGCTKAGGAALDKMTA---TCEKCGD	76
QY	66	YHFEY-----TECDSSGSRWR--VAIPNSAVDC-----	91
Db	77	GYFLFMGCGYKTTDGGPGEICTKAEGGLCTECKTANGLFKNPAATPEKGEICILCSDING	136
QY	92	-SGLPDPVRGKECTFS-----CASGEYLEMKNQVCSKCGEYTSLSGSIKFEDEW	139
Db	137	ADGYTGVANCAQCTKSDSNKGAATCTACQAGYKDF--QACSKC-DGT-----	181
QY	140	DELPAGESNIATFMDTVVGPSPDSRPDG-----CNNSSWIPRGNIESNRDDCTVSLI	191
Db	182	-----CLTCETSAQAQCTSCPECKYLKGDKSCVNNNGCTGNTYADPESGKC-----	226
QY	192	YAVHLKSGYVFFEYQYVDNNIFFEFFIQNDQCQEMDTTDDKKVKLTDNGEWSHVMKL	251
Db	227	-----LPCNTIDQAC-----TQC-EVDSTTKK-PRCTNCG-----	254
QY	252	SGTNILYWRTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCPKPGTFSNKPGSFNCQVCP	311
Db	255	-----GQKMKV-----TAIDG---TTTCVDANGCATSNVDGSH-----	284
QY	312	RNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFIHTPCDEEGKTQIMYKWIPE	371
Db	285	---FLNDGSTKILCSDDSSELEANKGT--PGC-----	312
QY	372	KICREDLTDAIRLPPSGEKKDCPPCNPFGFYNNGSSSCHPCPPGTFSDGTKECRPCPAGTE	431
Db	313	KTCKK-----NGAKPTCSECLDGYNSGNG-----GTVTCEACGANC-	349
QY	432	PALGFEYKWWNVLPGNMKTSCFNVNSKC-----DGMNGWEVAGDHIQSG----	476
Db	350	-----ATCTQAGNDKCTCKPKGFFMKNGPTGECVACDQAQGGIDGC	391
QY	477	-----AGSDNDYLILNLHIPGKPPPTSMTGATGSELGRITFV	514
Db	392	AECTKESTGPLKCTKCKPNRKPAQTSN-YTCTE---KTCENPTACGGTAGS-----	439
QY	515	FETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIIFKNATFTTWFQRTNQGO	574
Db	440	----CDA-----IVIDDQGTTHKHCYC-----GD	460



QY 575 DNRRFINDMVKIYSITATNAVGVASSCRACALGSEQSGSCVPCPPGHYIEKETNOCKE 634  
Db 461 SSQA-----PIDGL-----CASEAQKAGNTCA-----NGVCTQ 488  
QY 635 CPPDTYLSIHQVYGKEACIPCGPGS--KNNQDHSVCSYDCFFYHEKENQILHYDFSMLS 691  
Db 489 CTNNYFLYMGGCYSTQK---APGSPMCKTAGNTGIC-----TEAANN-----527  
QY 692 SVGSLMNGPSPFTSKGTYKFFHFNISLCGHEGKKMALCTNNITDFTVKEIVAGSDDYTN-L 750  
Db 528 -----RYFVVPGASNTDQSVLACSNPLGLT-----GTGDTAKAY 562  
QY 751 VGAFVQOSTIIPSESKGFRAALSOSIILADTFIGVTVETTLKNINIKEDMFPVPTSQIP 810  
Db 563 VGVEGCSQCTAP-----AALSDGGM-----582  
QY 811 DVHFFYKSTATTSCINGRSTAVKMRCNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESA 870  
Db 583 -----APAVCTSCDSSK-----KPNRDGSGCV---LCSVGCKKSCV-----615  
QY 871 EACPLCTEHDHFIEIGACKRGFOETLYVWNEPKWCIKGISLPEKKLATCETVDFWLKVGA 930  
Db 616 -----MDNICGECNSGF-----SLDNKCVSSGGANRSLGSLAGA 648  
QY 931 GVGAFTAVL-----LVALTCYFW 948  
Db 649 IAGISVAVVAVVGGVLVAFLCWWF 671  
RESULT 5  
T42215  
zonadhesin - mouse  
N:Alternate names: sperm-specific membrane protein  
C:Species: Mus musculus (house mouse)  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T42215  
R:Gao, Z.; Garbers, D.L.  
J. Biol. Chem. 273, 3415-3421, 1998  
A:Title: Species diversity in the structure of zonadhesin, a sperm-specific membrane protein  
A:Reference number: 222080; MUID:98123114; PMID:9452463  
A:Accession: T42215  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: mRNA  
A:Residues: 1-5376 <GAO>  
A:Cross-references: EMBL:U97068; NID:g3327420; PID:g3327421; PIDN:AAC26680.1  
C:Genetics:  
A:Gene: Zan  
A:Map position: 5  
C:Function:  
A:Description: functions in multiple cell adhesion processes  
A:Note: found exclusively on the apical region of the sperm head  
C:Keywords: cell adhesion  
Query Match 3.6%; Score 194; DB 2; Length 5376;  
Best Local Similarity 20.5%; Pred. No. 0.00053;  
Matches 144; Conservative 63; Mismatches 204; Indels 290; Gaps 38;  
QY 75 SSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNOVCSCGEGTYSLGSGI 134  
Db 3256 SRGCTQSTCTCPAGAIHCR-----NFKCPSGTCKNGDNGSSNCTEITLQCPTNS 3304  
QY 135 KFDEWDELPAFGFSNIATFMDTVVGPDSRDPDGCNNSWIPRGNYIESNRDDCTVSLIYAV 194  
Db 3305 QFT--DCLP-----SCVPSCSNRCEVTSVPSSCREGC-----3336  
QY 195 HLKKSQGVYFFEQYVDNNIFFEFFFIONDQCQEMDTTDDKVKVLTNDGEGWSHVMKLSGT 254  
Db 3337 -LCNHGFVFE-----DKCVPRTQCGCKDARGAIIPAG- 3368  
QY 255 NILYWRRTGILMGSKAVKPVLVKNITIEGVAYTSECFCKPFTFSKNKPGSFNCQ--VCPR 312  
Db 3369 -----KTWTSKGCTQSCACV-----EGNIQCQNFQCPP 3396

QY 313 NTYSEKAKECIRCKDDSQFSGSSECTE--RPPCTTKDYFOIHTPC-----DEEG 360  
Db 3397 ETY-----CKDNSE--GSSTCTKITLQCPAHTQ-----YTSCLPSCLPSCLDPEG 3439  
QY 361 KTOIMYKWIEPKI--CRE-----DLTDAI-RLPPSGEKKDCPP 395  
Db 3440 ---LCKDISPKVPSTCKEGCVQSGYVLNSDKVLRACEDCKDAQGALIPAGKTWTSPG 3495  
QY 396 CNPG-FYNNGSSSCH--PCPPPGTF---SDGTKECR---PCDAGTEPALGFYKWNVL 444  
Db 3496 CTQSCACMGGAVQCQSSQCPPTGYCKDNEDGNSNCAKITLQCPAHS-----LFTNCL 3547  
QY 445 PGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGSDNDYLILNLHIPGFKPPTSMTGAT 504  
Db 3548 PPCL-PSCLDP---DGL-----CKGASPKVPSTCKE 3574  
QY 505 GSELGRITVFVETLCSADCVLVFMVDINRKSTN--VVESWGGTKEKQAYTHIIFKNATFT 562  
Db 3575 G-----CICQSGYVL-----SNNKCLLRNRCGCKD--AHGALIPEDK---3609  
QY 563 FTWAFQRTNQODNRRFINDMVKIYSITATNAVGVASSCRACALGSEQSGSCVPCPPG 622  
Db 3610 -TWVSR-----GCTQSC-VCTGSGIOCLSS--QCPPG 3637  
QY 623 HYI---EKETNQCKECPDPTYLSIHQVYGKEACIPCGPGSKNNQDH-----SVCYS 670  
Db 3638 AYCKDNEDGSSNCARIPPOCPANSHYT---DCFPPCPPSCSDPEGHCEASGPRVLSTCRE 3694  
QY 671 DCF----FYHEKENQILHYDFSNLSSVSGSLM-NGPSFTSKG 706  
Db 3695 GOLCNPGFVLDRDKCVPVRECCKDAQGALIPSGKTTWTSPG 3735  
RESULT 6  
T23064  
hypothetical protein T22A3.8 - Caenorhabditis elegans (fragment)  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jul-2000  
C:Accession: T23064; T25096  
R:Barlow, K.  
submitted to the EMBL Data Library, October 1997  
A:Reference number: Z19669  
A:Accession: T23064  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-2823 <WIL>  
A:Cross-references: EMBL:AL008585; PIDN:CAA15432.1; GSPDB:GN00019; CESP:T22A3.8  
A:Experimental source: clone H10E24  
R:McMurray, A.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19980  
A:Accession: T25096  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-2823 <WI2>  
A:Cross-references: EMBL:Z81125; PIDN:CAB03385.1; GSPDB:GN00019; CESP:T22A3.8  
A:Experimental source: clone T22A3  
C:Genetics:  
A:Gene: CESP:T22A3.8  
A:Map position: 1  
A:Introns: 45/1; 282/2; 312/3; 416/2; 1255/3; 1329/3; 1418/3; 1776/2; 1988/2; 2760/2  
C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like  
Query Match 3.6%; Score 192.5; DB 2; Length 2823;  
Best Local Similarity 18.8%; Pred. No. 0.00031;  
Matches 205; Conservative 96; Mismatches 326; Indels 461; Gaps 59;  
QY 39 LAGCQAAWAGDLPPSSSSRPLPPC-----QEKDY-----HFEYTECDSSGS 78  
Db 819 LGACEQC---ECPSLDLNPNECISTELAVLGSVASNEDNYVCINCPLGYEGNKCEY---872  
QY 79 RWRVALPNSAVDCSG--LPDPVRGK--ECTF-----SCASGEYLEMKNOVCSCK- 123



Db 873 -----CSDGFFEDPLTGKIECTCNGNIDPMGIGNCDS-----ETGKCLKCI 914

QY 124 GEGTYSLGSIGIKFDEWDELPA GFSNIATFMDTVVGPDSRDPDCNNSSWIPRGNIESNR 183

Db 915 GHTTGDSCESCCKEHHW-----GNAQLHT-----CKPCGCHTQGA VNPQCSEENGE 959

QY 184 DDC TVSLIYAVHLKKG YVFF EYQYVDNNIFFE FFIQNDQCQEMDTTDDKWV KLT DNGEW 243

Db 960 CECKENYIGA-----QCDRC KENHGDV ENGCPACDCNDT 993

QY 244 GSHSVMLKSGTNILYWR TTGILMGSKAVKPV LVKNITIEGVAYTSEC FPCPKPGTFSN--- 300

Db 994 GS-----IGSDCDQ-----VSGQC-NCKQGVFGKQCD 1019

QY 301 --KPGSFN-----CQVCP RNTYSEKGAKECIRCKDDSQFSGSSECTERPPCT-----TKD 348

Db 1020 QCRPSYFNFTDAGCQFCHCNIY---GSIEDGKC---DQTTGKCEC RENVEGTMCEKCADG 1073

QY 349 YFQIHT-----PCDEEGKTQIMYKWI EPK-ICREDLTDAIRLPPSGEKKDCPPCNP GF 400

Db 1074 YFNITSGDGCEDCGDPTGSEDVSCNLVTGQCVC KPGVT-----GLK--CDSCLPNF 1123

QY 401 Y---NNGSSSCHPCP-PGTFSDGTKECRPCPAGTEPALGF EYKWWNVLPGNMKTSCFNVG 456

Db 1124 YGLTSEGCTECP PAPGVCDPIDGSCVCP PNT-----V GEMCENCIT-- 1167

QY 457 NSKCDGMNGWEVAGDHIQSGAGSDNDYLILNLH IPIPKPPTSMTGATGSELGRITFVFE 516

Db 1168 -----NAW-----DYHPLN-----GCK----- 1179

QY 517 TLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAY-----THIIFKNATFTT WAFQ 568

Db 1180 -LC--DC-----SDIGSDGGM CNTFTTGQCKCKAA YVGLKCDLCTHGFNFPT----- 1223

QY 569 RTNOGQDNRRFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGHYIEKE 628

Db 1224 -----CEPC--GCNAAGTDPLQCKDGQCL-- 1245

QY 629 TNQCKECPDPTYLSIHQVYGRACIPCGPG-----SKNNQDHSVCYSDCF----- 673

Db 1246 CNEIGEC P-----CKKNVHG TK-CDQC GEGTFSLDSSNLKGCTECF--CFNRTSNCEQSD 1297

QY 674 -----FYHEKEN-QILHYDFS NLSS-----V GSLMNGPS 701

Db 1298 LVWQMYAEDRRAVFQEPWEFYTKKHNNILLREKPSHFNSYPTDATPLYWLPSTMLGDR 1357

QY 702 FTS-----KGTKYF-HFFNISLCGHEGKKMALCTNNITDFTVK 738

Db 1358 TASYNGFLRFKIWNEDNRRRGLHGIRPDQOYFRHFPQV IIFGNRIELEHIPMEINDDGIY 1417

QY 739 EI-----VAGSDDYT-----NLVGA FVCQSTIIPSESKGFRAALSSQSIIL 779

Db 1418 KIRLHESEWRVRHSPELTLTRKQMMVALQDTQGIYIRGT YTPARGDAINIQEVSLDVAV 1477

QY 780 ADTFIGVTVETTLKNINIKEDM-----FPVPTSQIPDVHFFYKSS----- 819

Db 1478 PESKIVAGLSTT-KAIGVEKLGCPQGYTGLSCQNPEVGYRKKHREYLNQADDIALIGW 1536

QY 820 TATTSCINGRSTAVKMR CNPTKSGAGVISVPSKCPAGT-----CDGCTFYFLWESAE--- 871

Db 1537 SEPCSC-HGHS-----QTCNPD-----TSVCTDCEHNTFGDFCEHCLPGYIGDAREGGAN 1585

QY 872 -----ACPL-----CTEHDFHEIEGACKRGFOETLYVWN EPKWC IKG-ISLPEKK 915

Db 1586 ACTKCACPLVENSFSDSCVAVDHGRGYVCNSCKPG-----YTGQYCETCVAGYGD PQHI 1640

QY 916 LATCETVD 923

Db 1641 GGTCSPCD 1648

RESULT 7

F87908

protein T22A3.8 [imported] - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-Aug-2001

C;Accession: F87908; E87908

R;anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating bio

A;Reference number: A75000; MUID:99069613; PMID:9851916

A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_-

A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;

A;Accession: F87908

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-2823 <STO>

A;Cross-references: GB:chr\_I; PIDN:CAA15432.1; PID:g3924779; GSPDB:GN00019; CESP:T22A

A;Accession: E87908

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-2823 <ST2>

A;Cross-references: GB:chr\_I; PIDN:CAB03385.1; PID:g3924881; GSPDB:GN00019; CESP:T22A

C;Genetics:

A;Gene: T22A3.8

A;Map position: 1

C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-lik

Query Match 3.6%; Score 192.5; DB 2; Length 2823;

Best Local Similarity 18.8%; Pred. No. 0.00031;

Matches 205; Conservative 96; Mismatches 326; Indels 461; Gaps 59;

QY 39 LAGQAAWAGDLPSSSRPLPPC-----QEKDY-----HFEYTECDSSGS 78

Db 819 LGACEQC---ECPSLDLNPNPECISTELAVLGSVASNEDNYVCINCP LGYEGNKCEY--- 872

QY 79 RWRVAIPNSAVDCSG--LPDPVRGK--ECTF-----SCASGEYLEMKNQVCSKC- 123

Db 873 -----CSDGFFEDPLTGKIECTCNGNIDPMGIGNCDS-----ETGKCLKCI 914

QY 124 GEGTYSLGSIGIKFDEWDELPA GFSNIATFMDTVVGPDSRDPDCNNSSWIPRGNIESNR 183

Db 915 GHTTGDSCESCCKEHHW-----GNAQLHT-----CKPCGCHTQGA VNPQCSEENGE 959

QY 184 DDC TVSLIYAVHLKKG YVFF EYQYVDNNIFFE FFIQNDQCQEMDTTDDKWV KLT DNGEW 243

Db 960 CECKENYIGA-----QCDRC KENHGDV ENGCPACDCNDT 993

QY 244 GSHSVMLKSGTNILYWR TTGILMGSKAVKPV LVKNITIEGVAYTSEC FPCPKPGTFSN--- 300

Db 994 GS-----IGSDCDQ-----VSGQC-NCKQGVFGKQCD 1019

QY 301 --KPGSFN-----CQVCP RNTYSEKGAKECIRCKDDSQFSGSSECTERPPCT-----TKD 348

Db 1020 QCRPSYFNFTDAGCQFCHCNIY---GSIEDGKC---DQTTGKCEC RENVEGTMCEKCADG 1073

QY 349 YFQIHT-----PCDEEGKTQIMYKWI EPK-ICREDLTDAIRLPPSGEKKDCPPCNP GF 400

Db 1074 YFNITSGDGCEDCGDPTGSEDVSCNLVTGQCVC KPGVT-----GLK--CDSCLPNF 1123

QY 401 Y---NNGSSSCHPCP-PGTFSDGTKECRPCPAGTEPALGF EYKWWNVLPGNMKTSCFNVG 456

Db 1124 YGLTSEGCTECP PAPGVCDPIDGSCVCP PNT-----V GEMCENCIT-- 1167

QY 457 NSKCDGMNGWEVAGDHIQSGAGSDNDYLILNLH IPIPKPPTSMTGATGSELGRITFVFE 516

Db 1168 -----NAW-----DYHPLN-----GCK----- 1179

QY 517 TLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAY-----THIIFKNATFTT WAFQ 568

Db 1180 -LC--DC-----SDIGSDGGM CNTFTTGQCKCKAA YVGLKCDLCTHGFNFPT----- 1223

QY 569 RTNOGQDNRRFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGHYIEKE 628

Db 1224 -----CEPC--GCNAAGTDPLQCKDGQCL-- 1245

3

```
QY 629 TNQCKECPDITYLSIHQVYGKEACIPCGPG-----SKNQDHSVCYSDCF----- 673
Db 1246 CNEIGECP-----CKKNVHGTK-CDQCGEGTFLDSSNLKGCTECF--CFNRTSNCEQSD 1297
QY 674 -----FYHEKEN-QILHYDFSNLSS-----VGSMLMNGPS 701
Db 1298 LVWQOMYAEARRAVFQEPWEFYTKKHNNLLREKPSHENSYPDTATPLYWPLPSTMLGDR 1357
QY 702 FTS-----VAGSDDYT-----KGTKYF-HFFNISLCGHEGKKMALCTNNITDFTVK 738
Db 1358 TASYNGFLRFKIWNEDNRRLHGIIRPDQQYFRHFPPQVIIFGNRRIELEHIPMEINDDGIY 1417
QY 739 EI-----VAGSDDYT-----NLVGAFCVQSTIIPSESKGFRAALSSQSIL 779
Db 1418 KIRLHESEWRVRHSPELTLTRKQMMVALQDTQGIYIRGTITYTPARGDAINIQEVSLDVAV 1477
QY 780 ADTFIGVTVETTLKNINIKEDM-----FPVPTSQIPDVHFFYKSS----- 819
Db 1478 PESKIVAGLSTT-KAIGVEKCLGCPQGYTGLSCQNPEVGYRKKHREYLNQADDIALIGW 1536
QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGT-----CDGCTFYFLWESAE----- 871
Db 1537 SEPCSC-HGHS---QTCNPD-----TSVCTDCEHNTFGDFCEHCLPGYIGDAREGGAN 1585
QY 872 -----ACPL-----CTEHDFFEIEGACKRGFQETLYVWNEPKWCIKG-ISLPEKK 915
Db 1586 ACTKACPLVENSFSDSCVAVDHGRGYVCNSCKPG-----YTGQYCETCVAGYGGDPQHI 1640
QY 916 LATCETVD 923
Db 1641 GGTCSPCD 1648

RESULT 8
T43291
laminin alpha chain - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C:Accession: T43291
R:Zhu, X.; Kao, G.; Joh, K.; Hall, D.H.; Wadsworth, V.K.; Hutter, H.; Vogel, B.E.; Huang
submitted to the EMBL Data Library, June 1998
A:Description: Expression, function and evolution of laminin alpha chains.
A:Reference number: 222397
A:Accession: T43291
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3102 <2HU>
A:Cross-references: EMBL:AF074902; PIDN:AAC26793.1
C:Genetics:
A:Map position: 1
A>Note: lamal/2
C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h

Query Match 3.6%; Score 192.5; DB 2; Length 3102;
Best Local Similarity 18.8%; Pred. No. 0.00035;
Matches 205; Conservative 96; Mismatches 326; Indels 461; Gaps 59;

QY 39 LAGCQAAWAGDLPSSSSRPLPPC-----QEKDY-----HFEYTECDSSGS 78
Db 819 LGACEQC---ECPSLDLNPNPECISTELAVLGSVASNEDNYVCINCPGLYEGNKCEY--- 872
QY 79 RWRVAIPNSAVDCSG--LPDPVRGK--ECTF-----SCASGEYLEMKNQVCSKC- 123
Db 873 -----CSDGFFEDPLTGKICTECTCNGNIDPMGIGNCDS-----ETGKCLKCI 914
QY 124 GEGTYSLGSCTKFDWEDELPAAGFSNIATFMDTVVGPDSRDPDGCNNSSWIPRGNYIESNR 183
Db 915 GHTTGDSCECKEHHW-----GNAQLHT-----CKPCGCHTQGA VNPQCSENGE 959
QY 184 DDCTVSLIYAVHLKSGYVFFEYQYVDNNIFFEFFFIONDQCQEMDTTDDKWKLTNDGEW 243
Db 960 CECKENYIGA-----QCDRCKENHGDVNGCPACDCNDT 993
```

```
QY 244 GSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFCPKPGTFSN--- 300
Db 994 GS-----IGSDCDQ-----VSGQC-NCKQGVFGKQCD 1019
QY 301 --KPGSFN-----CQVCPRNTYSEKGAKEIRCKDDSQFSGSSECTERPPCT-----TKD 348
Db 1020 QCRPSYFNFTDAGCQFCHCNIY---GSIEDGKC---DOTTGKCECENVEGTMEKCADG 1073
QY 349 YFQIHT-----PCDEEGKTQIMYKWIIEPK-ICREDLTDAIRLPPSGEKKDCPPCNPGF 400
Db 1074 YFNITSGDGCEDCGCDPTGSEDVSCNLVLTGQCVCCKPGVT-----GLK--CDSCLPNF 1123
QY 401 Y---NNGSSSCHPCP-PGTFSDGTKECRPCPAGTEPALGFYKWWNVLPGNMKTSCFNVG 456
Db 1124 YGLTSEGCTECEPCPAPGQVCDPIDGSCVCPNT-----VGENCENCCTT-- 1167
QY 457 NSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIFGFKPPTSMTGATGSELGRITFEVFE 516
Db 1168 -----NAW-----DYHPLN-----CCK----- 1179
QY 517 TLCSADCVLYFMVDINRKSTNVVSWGGTKEKQAY-----THIIFKNATFTTWAFAQ 568
Db 1180 -LC--DC-----SDIGSDGMCNTFTGQCKCKAAYVGLKCDLCTHGFEFNT----- 1223
QY 569 RTNQGDNRFRINDMVKIYSTATNAVDGVASSCRACALGSESGSSCVPCPPGPHYIEKE 628
Db 1224 -----CEPC--GCNAAGTDPLQCKDGQCL--- 1245
QY 629 TNQCKECPDITYLSIHQVYGKEACIPCGPG-----SKNQDHSVCYSDCF----- 673
Db 1246 CNEIGECP-----CKKNVHGTK-CDQCGEGTFLDSSNLKGCTECF--CFNRTSNCEQSD 1297
QY 674 -----FYHEKEN-QILHYDFSNLSS-----VGSMLMNGPS 701
Db 1298 LVWQOMYAEARRAVFQEPWEFYTKKHNNLLREKPSHENSYPDTATPLYWPLPSTMLGDR 1357
QY 702 FTS-----KGTKYF-HFFNISLCGHEGKKMALCTNNITDFTVK 738
Db 1358 TASYNGFLRFKIWNEDNRRLHGIIRPDQQYFRHFPPQVIIFGNRRIELEHIPMEINDDGIY 1417
QY 739 EI-----VAGSDDYT-----NLVGAFCVQSTIIPSESKGFRAALSSQSIL 779
Db 1418 KIRLHESEWRVRHSPELTLTRKQMMVALQDTQGIYIRGTITYTPARGDAINIQEVSLDVAV 1477
QY 780 ADTFIGVTVETTLKNINIKEDM-----FPVPTSQIPDVHFFYKSS----- 819
Db 1478 PESKIVAGLSTT-KAIGVEKCLGCPQGYTGLSCQNPEVGYRKKHREYLNQADDIALIGW 1536
QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGT-----CDGCTFYFLWESAE----- 871
Db 1537 SEPCSC-HGHS---QTCNPD-----TSVCTDCEHNTFGDFCEHCLPGYIGDAREGGAN 1585
QY 872 -----ACPL-----CTEHDFFEIEGACKRGFQETLYVWNEPKWCIKG-ISLPEKK 915
Db 1586 ACTKACPLVENSFSDSCVAVDHGRGYVCNSCKPG-----YTGQYCETCVAGYGGDPQHI 1640
QY 916 LATCETVD 923
Db 1641 GGTCSPCD 1648
```

RESULT 9  
T25933  
hypothetical protein W02C12.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T25933  
R:Murray, J.; Wohldmann, P.  
submitted to the EMBL Data Library, December 1996  
A:Description: The sequence of C. elegans cosmid W02C12.  
A:Reference number: 220112  
A:Accession: T25933





C;Species: Giardia lamblia  
C;Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 01-Dec-2000  
C;Accession: A42125; B42125; S00530; S48056  
R;Adam, R.D.; Yang, Y.M.; Nash, T.E.  
Mol. Cell. Biol. 12, 1194-1201, 1992  
A;Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170 gene in Giardia lamblia  
A;Reference number: A42125; MUID:92186850; PMID:1545800  
A;Accession: A42125  
A;Molecule type: DNA  
A;Residues: 1-98 <ADA1>  
A;Cross-references: GB:M83937; NID:g159124  
A;Experimental source: trophozoite  
A;Note: sequence extracted from NCBI backbone (NCBIN:88421, NCBIP:88427); this ORF is not in the authors report but do not show 19 tandem repeats of the sequence of residues 1-98  
A;Accession: B42125  
A;Molecule type: DNA  
A;Residues: 1269-1766 <ADA2>  
A;Cross-references: GB:M83933; NID:g159122  
A;Note: sequence extracted from NCBI backbone (NCBIN:88424, NCBIP:88431); this ORF is not in the authors report but do not show 19 tandem repeats of the sequence of residues 1-98  
R;Adam, R.D.; Aggarwal, A.; Lal, A.A.; de la Cruz, V.F.; McCutchan, T.; Nash, T.E.  
J. Exp. Med. 167, 109-118, 1988  
A;Title: Antigenic variation of a cysteine-rich protein in Giardia lamblia.  
A;Reference number: S00530; MUID:88089405; PMID:3335828  
A;Accession: S00530  
A;Molecule type: DNA  
A;Residues: 1154-1409, 'A', 1411-1420, 'K', 1422-1425, 'R', 1427-1481 <ADA3>  
A;Cross-references: EMBL:X06741; NID:g9355; PID:g929603  
R;Yang, Y.; Adam, R.D.  
Nucleic Acids Res. 22, 2102-2108, 1994  
A;Title: Allele-specific expression of a variant-specific surface protein (VSP) of Giardia lamblia  
A;Reference number: S48056; MUID:94301794; PMID:8029018  
A;Accession: S48056  
A;Molecule type: DNA  
A;Residues: 1-56 <YAN>  
A;Cross-references: EMBL:L25059  
A;Experimental source: trophozoites WBA6  
A;Note: the source is designated as Giardia intestinalis  
C;Comment: This translation was produced by PIR staff from information provided by the authors  
C;Genetics:  
A;Gene: VSPA6  
C;Keywords: surface antigen; tandem repeat

Query Match 3.4%; Score 181; DB 2; Length 1766;  
Best Local Similarity 20.6%; Pred. No. 0.001;  
Matches 233; Conservative 107; Mismatches 432; Indels 361; Gaps 67;

QY 20 APRRGRSPWPSPAWICWALAGCOAAWAGDLPSSSSRPLPPCOBKDYHF-----EYT 71  
Db 693 APYLKKTNPSTGTGTCVSAV-DCQGS-AGYTTDDSVSDAKECKKCNAPCTACAGTADKCT 750  
QY 72 ECDSSGSRWRVAIPN-----SAVDCSG-----LPDPVRGKECTFSCASGEYL 113  
Db 751 KCDANGAAPYLKKTNPSTGTGTCVSAVDCQGSAGYTTDDSVSDAKECKKCNAPCTACAGT 810  
QY 114 EMKNQVCSKC---GEGTY-----SLSGIKFDEWD-ELPAGFSNIATFMDTVVGPSPDSR 163  
Db 811 ADK---CTKCDANGAAPYLKKTNPSTGTGTCVSAVDCQGSAGY-----YTDDSV---SDAK 860  
QY 164 -----PDGCNNSWIIPRGNYI-ESNRDDCTVSLIYAVHLKKS-GYVFFEY 206  
Db 861 ECKKCNAPCTACAGTADKCTKCDANGAAPYLKKTNPSTGTGTCVSAVDCQGSAGY----- 915  
QY 207 QYVDNNI--FPEFFIQNDQCQEMDTTDDKWKVLTNDGEWGSVMLK-----SGTNIL 257  
Db 916 -YTDDSVSDAKECKKCNAPCTACAGTADKCTKCDANGA---APYLKKTNPSTGTGTCV- 969  
QY 258 YWRTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCPKPGTFSNKPGSFNCQVCPRN---- 313  
Db 970 ---SAVDCQGSAGY-----YTDDSVSDAKECKKCNAPCTACAGTADKCTKCDANGAAP 1019  
QY 314 ----TYSEKGAKECIRCKD-----DSQFSGSSECTE-RPPCT----TKDYFIHTP 355  
Db 1020 YLKKTNPSDPTGTGTCVSAVDCQGSAGYTTDDSVSDAKECKKCNAPCTACAGTADKC---TK 1076

QY 356 CDEEGKTQIMYKWIPIKICREDLTDAILRPPSGEKKDCPPC--NPGFYNNNGSSS-----CH 409  
Db 1077 CDANGAA-----PYLKKTNPSD-----PTGTCVSAVDCQGSAGYTTDDSVSDAKECK 1123  
QY 410 PC-PPGTFSDGTKE-CRPPCPA-GTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGW 466  
Db 1124 KGNAPCTACAGTADKCTKCDANGAAPYL---KKTN--PSDPTGTC--VSAVDCQGSAGY 1175  
QY 467 EVAGDHIQSG-----AGGSDNDYLILNLHPGFKP-----PTSMTGATGSE 507  
Db 1176 -YTDDSVSDAKECKKCNAPCTACAGTADK---CTKCDANGAAPYLKKTNPSTPTG----- 1226  
QY 508 LGRITFVFETLCSADCVLXFVMDIN-----RKSTNVVSVGGTKEKQAYTHIIFKNATFT 562  
Db 1227 ---TCVSAVDCQGSAGYTTDDSVSDAKECKKCNAPCTACAGTADKCTKCD---ANGAAP 1279  
QY 563 FTWAFQRTNQGDNRFRINDVMVKIYSITATNAV-----GVASSCRACALGS 609  
Db 1280 Y---LKKTNPSDPTG-----TCVSAVDCQGSAGYTTDDSVSDAKECKKCAEGQ 1324  
QY 610 E--QSGSSCVPCPPGHYIEKETNQ-CKEC---PPDTVLSIHQVYGEACIPCGPGSKNN 662  
Db 1325 KPNTAGTQCFSCSDANCERCQDNDVCARCSTGAPPEN-----GK--CPAATPG---- 1370  
QY 663 QDHSVCYSDCFFYHEKENQILHYDFSLSVSGSLMNGPSTSGTKYFHEFN-ISLCGHE 721  
Db 1371 -----CHSSCDGCTE-----NMTNQADKCTGCKEGRYLKPESAAGQS 1408  
QY 722 GKKNAL--CTNNITDFTVKE-----IVAGSDDYTNLVGAFVC-----QSTIIPS 763  
Db 1409 GTCLTAECTSDTTHFTKEKAGDSKGMCLSCSDATHGITCKKCALKTLSGEAEESTVVC 1468  
QY 764 ESKGFRAALSSQSII-----LADTFIGTVV---ETTLKNINIKEDMFVPTSQIPDV 812  
Db 1469 ECTDKRLTPSGNACLDNCPAGTYADNINGVSVASCASHATCAECNGDANAASC-TACYPGY 1527  
QY 813 HFFYKSSTA-----TTSCINGRSTA---VKMRCNPTKSGAGVI----- 847  
Db 1528 SLLYGSSTAGTCVKECTGAFITNCADGQCTANVGGAKYCAQCKDGYAPIDGICTTVAAG 1587  
QY 848 -----SVPSKCP-----AGTCDG-----CTFYFLWESAECPLCTEHDFFH 882  
Db 1588 RDASVCTAADGKCTKAGCYTLMGGCYGVAKLPGKSVCTL-----ASNGKCTMCAANQA 1643  
QY 883 EIEGACKRGFQETLYVWNEPKWCTKISLPEKKLATCETVDFWLKVAGVGAF 935  
Db 1644 PVEKCP-----CSEGCACNDSNACTECLPGYK-GAGDKCF 1681

RESULT 12  
T23433

hypothetical protein K08C7.3 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000

C;Accession: T23433

R;Berks, M.

submitted to the EMBL Data Library, March 1996

A;Reference number: Z19740

A;Accession: T23433

A;Status: preliminary; translated from GB/EMBL/DBD

A;Molecule type: DNA

A;Residues: 1-3672 <WIL>

A;Cross-references: EMBL:Z70286; PIDN:CAA94293.1; GSPDB:GN00022; CESP:K08C7.3

A;Experimental source: clone K08C7

C;Genetics:

A;Gene: CESP:K08C7.3

A;Map position: 4

A;Introns: 66/1; 284/3; 563/1; 1187/3; 1248/3; 1300/1; 1460/1; 1623/3; 2361/3; 2988/3

C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like

Query Match 3.3%; Score 178.5; DB 2; Length 3672;

Best Local Similarity 22.6%; Pred. No. 0.0036;



Matches 116; Conservative 54; Mismatches 203; Indels 141; Gaps 30;

QY	279	ITIEGVAYTSECFFCKP--GTFSNKPGSF--NCQVCPRNTYSEKGAKECIRCKDDSQFSG	334
Db	1413	LSCDCVAQGSSEFQCEQYGGQCKKPGVIGRRRCERCAPGYN--FPECIKC-----QCNA	1465
QY	335	SSECTERP-PCTTKDYFOIHTPCDEEGKTQIMYKWIPIK-----CREDLTDAIRL---	384
Db	1466	GQQCDERTGQCFCPHVEGQT-CDRCVSNAFY---DPLIGCQKCGCHPQGSSEGNLVCD	1521
QY	385	PPSGE-----KKDCPPCNPFGYNNGSSSCHPCP---PGT---FSDGK-----	421
Db	1522	PESGQCLCRESMGGRQCDRLAGFY--GFPHCYGCSNCRAGTTEEICDATNAOCKCKENV	1579
QY	422	---ECRPCPAGTEPALGFYKWWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAG	478
Db	1580	YGRCEACKAGT-----FDLSAENPL-GCVNCFCFGVTDSCRSSMYPVTIMSDMSSFLT	1633
QY	479	GSDN-----DYLILNLHIPG-----FKPPTSMTGATGSELGRITFVFETL-----C	519
Db	1634	TDDNGMVNDKDDTVITYTSEETSPNSVYFNVPIEKKDYTTSYGLKLTFTFKLSTVPRGRKSM	1693
QY	520	SADCVLYFMVDINRKSTNV-VESWGG---TKEQAYTHIIFKNATFTFTTWFQRTNQGD	575
Db	1694	NAD-----ADVRLTGANMTIEYWASEQPTNPEQFT-VKCKLVPENFLTAECKTVTREE	1746
QY	576	NRRFINDMVKI-----YSITATNAVDGV--ASSCRACALGSEQSGS	614
Db	1747	LMKVLHSLQNTLTKASYFDHPKTSTLYEFGLEISEPNGVDSVIKASSVEQOCQCPAPYTG	1806
QY	615	SCVPCPPGHYIEKETNQCKECPDPTLYLSIHQYVKE---ACIPC---GPGSKNQDHSVC	668
Db	1807	SCQLCASGY-----HRVQSGSFLGACVPCECNHGSATCDPDTGIC	1846
QY	669	YSDCFFYHEKENQILHYDFSNLSSVGSGLMNGPSF	702
Db	1847	-TDC-----EHNNTNGDHCEFCNEGHYGNATNGSPY	1875

RESULT 13

T37316  
probable laminin alpha chain - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Jan-2000  
C:Accession: T37316  
R:Joh, K.; Zhu, K.; Hedgecock, E.M.; Inoue, T.; Hori, K.  
submitted to the EMBL Data Library, August 1998  
A:Description: Laminin alpha chain gene in the nematode C. elegans.  
A:Reference number: Z21681  
A:Accession: T37316  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3704 <JOH>  
A:Cross-references: EMBL:AB016806; PIDN:BAA32347.1  
A:Experimental source: strain N2  
C:Genetics:  
A:Gene: epi-1  
A:Map position: IV  
A:Introns: 66/1; 284/3; 563/1; 1187/3; 1248/3; 1300/1; 1460/1; 1623/3; 2361/3; 2988/3;  
C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like

Qy	279	ITIEGVAYTSECFPCPKP--GTFSNKPGSF--NCQVCPRNTYSEKGAKECIRCKDDSQFSG	334
Db	1413	LSCDCVAQSGSFQCEQYGGQCKKPGVIGRRERCAPGYN---FPECIKC-----QCNA	1465
Qy	335	SSECTERP-PCTTKDYFIHTPCDEEGKTQIMYKWIIEPKI-----CREDLTDAILR---	384
Db	1466	GQQCDERTGQCFCPPHVEGQT-CDRCVSNAFGY---DPLIGCQKCGCHPQSGEGGNLVCD	1521

```

QY 385 PPSGE-----KKDCPPCPNGFYNNSSSCHPCP---PGT---FSDGTK----- 421
      |||:      | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1522 PESGQCLRESNGRQCDRLAGFY--GPHCYGCSNRRAGTTEEICDATNAQCKCKENV 1579
      | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 422 ---ECRPCAGTEPALGFYKWNVLPGNMKTSFCFVNGSKCDGMNGWEVAGDHIQSGAG 478
      | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1580 YGGRCEACKAGT-----FDLSAENPL-GCVNCFCFGVTDSCRSSMYPVTIMSVDMSSFLT 1633
      | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 479 GSDN-----DYLILNLHIPG-----FKPPTSMTGATGSELGRITVFETL-----C 519
      ||      | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1634 TDDNGMVDNKDDIVITYTSEETSPNSVYFNVPIEKKDYTTSYGLKLTFTKLTSTVPRGGRKSM 1693
      | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 520 SADCVLVFMVDINRKSTNV-VESWGG---TKEKQAYTHIIFKNATFTFTWAFQRTNQGD 575
      :||      | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1694 NAD-----ADVRLTGANMTIEYWASEQPTNPEEQFT-VKCKLVPENFLTAEGKTVTREE 1746
      | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 576 NRRFINDMVKI-----YSITATNAVQGV--ASSCRACALGSEQSGS 614
      : : : | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1747 LMKVLHLSLQNTILKASYFDHPKSTLYEFGLEISEPNGVDSVIKASSVEQCQCPAPYTG 1806
      | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 615 SCVPCPPGHVIEKETNQCKECPDPTLYLSIHQVYGKE---ACIPC---GPGSKNNQDHSVC 668
      || | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1807 SCQLCASGY-----HRVQSGSFLGACVPCECNHGSATCDPDTGIC 1846
      | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 669 YSDCFYHEKENQILHYDFSLSVSGSLMNGPSF 702
      :|| | | | | | | | | | | | | | | | | | | | | | | | |
Db 1847 -TDC-----EHTNGDHCFCFCNEGHYGNATNGSPY 1875
      | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
MMMSB2
laminin gamma-1 chain precursor - mouse
N:Alternate names: laminin chain B2
C:Species: Mus musculus (house mouse)
C:Date: 28-Feb-1986 #sequence_revision 30-Jun-1991 #text_change 10-Dec-1999
C:Accession: A28469; A27729; A28082; S02680; S05327; S02037; A02870; S13544; S14552
R:Sasaki, M.; Yamada, Y.
J. Biol. Chem. 262, 17111-17117, 1987
A:Title: The laminin B2 chain has a multidomain structure homologous to the B1 chain.
A:Reference number: A28469; MUID:88059118; PMID:3680290
A:Accession: A28469
A:Molecule type: mRNA
A:Residues: 1-1607 <SAS>
A:Cross-references: EMBL:J03484; NID:g198694; PIDN:AAA39405.1; PID:g293688
R:Durkin, M.E.; Barlos, B.B.; Liu, S.H.; Phillips, S.L.; Chung, A.E.
Biochemistry 27, 5198-5204, 1988
A:Title: Primary structure of the mouse laminin B2 chain and comparison with laminin
A:Reference number: A27729; MUID:89000737; PMID:3167041
A:Accession: A27729
A:Molecule type: mRNA
A:Residues: 1-263,'D',265-336,'C',338-446,'PS',449-661,'S',663-885,887-1155,1157-1433
A:Cross-references: EMBL:J02930; NID:g198702; PIDN:AAA39408.1; PID:g293691
A:Note: The authors translated the codon TAT for residue 544 as Asp and GCG for resid
R:Ogawa, K.; Burbello, P.D.; Sasaki, M.; Yamada, Y.
J. Biol. Chem. 263, 8384-8389, 1988
A:Title: The laminin B2 chain promoter contains unique repeat sequences and is active
A:Reference number: A28082; MUID:88228071; PMID:2836421
A:Accession: A28082
A:Molecule type: DNA
A:Residues: 1-215,'A',217-239 <OGA>
A:Cross-references: EMBL:J03749; NID:g198704; PIDN:AAA39409.1; PID:g554184
R:Fujiwara, S.; Shinkai, H.; Deutzmann, R.; Paulsson, M.; Timpl, R.
Biochem. J. 252, 453-461, 1988
A:Title: Structure and distribution of N-linked oligosaccharide chains on various dom
A:Reference number: S02678; MUID:88326259; PMID:2458101
A:Accession: S02680
A:Molecule type: protein
A:Residues: 227-238 <FUJ>
R:Hartl, L.; Oberbaeumer, I.; Deutzmann, R.
Eur. J. Biochem. 173, 629-635, 1988
A:Title: The N terminus of laminin A chain is homologous to the B chains.
A:Reference number: S00624; MUID:88225080; PMID:3267223
A:Accession: S05327

```

A;Molecule type: protein  
A;Residues: 227-238;387-393,'F',395-405;881-912;1022-1034 <HAR>  
R;Deutzmann, R.; Huber, J.; Schmetz, K.A.; Oberbaeumer, I.; Hartl, L.  
Eur. J. Biochem. 177, 35-45, 1988  
A;Title: Structural study of long arm fragments of laminin. Evidence for repetitive C-ter  
A;Reference number: S01790; MUID:89030693; PMID:3181157  
A;Accession: S02037  
A;Molecule type: protein  
A;Residues: 1362-1377,'X',1379-1392,'X',1394-1406 <DEU>  
R;Barlow, D.P.; Green, N.M.; Kurkinen, M.; Hogan, B.L.M.  
EMBO J. 3, 2355-2362, 1984  
A;Title: Sequencing of laminin B chain cDNAs reveals C-terminal regions of coiled-coil a  
A;Reference number: A02870; MUID:85051302; PMID:6209134  
A;Accession: A02870  
A;Molecule type: mRNA  
A;Residues: 1391-1474,'K',1476-1575,'N',1577-1607 <BAR>  
A;Cross-references: EMBL:X05211; NID:952862; PIDN:CAA28838.1; PID:g817975  
R;Paulsson, M.; Deutzmann, R.; Timpl, R.; Dalzoppo, D.; Odermatt, E.; Engel, J.  
EMBO J. 4, 309-316, 1985  
A;Title: Evidence for coiled-coil alpha-helical regions in the long arm of laminin.  
A;Reference number: S13543; MUID:85257455; PMID:3848400  
A;Accession: S13544  
A;Molecule type: protein  
A;Residues: 1506-1523,'X',1525 <PAU>  
R;Olson, D.; Nagayoshi, T.; Fazio, M.; Peltonen, J.; Jaakkola, S.; Sanborn, D.; Sasaki,  
Lab. Invest. 60, 772-782, 1989  
A;Title: Human laminin: cloning and sequence analysis of cDNAs encoding A, B1 and B2 cha  
A;Reference number: A34961; MUID:89280632; PMID:2733383  
A;Accession: S14552  
A;Molecule type: protein  
A;Residues: 881-912;1022-1034;1364-1377;1379-1392;1394-1409;1506-1525;1593-1606 <OLS>  
C;Genetics:  
A;Gene: Lamb-2  
A;Map position: 1  
A;Introns: 138/1; 239/3  
C;Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin  
C;Function:  
A;Description: Interact with cells and with other basement membrane proteins to promote  
C;Superfamily: laminin beta-1 chain; laminin-type EGF-like homology  
C;Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular  
F;1-33/Domain: signal sequence #status predicted <SIG>  
F;34-1607/Product: laminin gamma-1 chain #status predicted <MAT>  
F;34-283/Domain: VI <DOM6>  
F;284-502/Domain: V <DOM5>  
F;284-337/Domain: laminin-type EGF-like homology #status atypical <LE01>  
F;340-393/Domain: laminin-type EGF-like homology <LE02>  
F;396-440/Domain: laminin-type EGF-like homology <LE03>  
F;443-490/Domain: laminin-type EGF-like homology <LE04>  
F;493-502/Domain: laminin-type EGF-like homology #status atypical <LE05>  
F;503-687/Domain: IV <DOM4>  
F;688-1032/Domain: III <DOM3>  
F;688-719/Domain: laminin-type EGF-like homology #status atypical <LE06>  
F;722-768/Domain: laminin-type EGF-like homology <LE07>  
F;771-823/Domain: laminin-type EGF-like homology <LE08>  
F;826-879/Domain: laminin-type EGF-like homology <LE09>  
F;882-930/Domain: laminin-type EGF-like homology <LE10>  
F;933-978/Domain: laminin-type EGF-like homology <LE11>  
F;981-1026/Domain: laminin-type EGF-like homology <LE12>  
F;1033-1607/Domain: II/I <DOM2>  
F;1033-1607/Region: heptad repeats  
F;38-48/Disulfide bonds: #status predicted  
F;58,132,574,648,1020,1105,1159,1173,1203,1221,1239,1437/Binding site: carbohydrate (Asn  
F;1029,1032/Disulfide bonds: interchain #status predicted  
F;1378,1393/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F;1598/Disulfide bonds: interchain (to chain B1) #status experimental

Query Match 3.3%; Score 177.5; DB 1; Length 1607;  
Best Local Similarity 19.3%; Pred. No. 0.0016;  
Matches 215; Conservative 115; Mismatches 437; Indels 347; Gaps 57;

QY 13 GWRPAEAPR-RGRSPWPSPAWICWALAGQAAGDLPSSSSRPLPPCQEKDYHFEYT 71  
Db 3 GGGRAALALQPRGR-----LWPLLAVALAAGVAGCVRAAMD----- 36

QY 72 ECDSSGSRWRVAIP---NSAVD-----CSGLPDPVRGKECTFSCASGEYLEMKNQVC 120  
Db 37 ECADEGGRPQRCMPEFVNAFNVTVATNTCGTPE---EYCVQGTG-----VTKSC 87  
QY 121 SKCGEGTYSLSGSIKF--DEWDELPAFNSIATEFMDTVVGPSPD----- 161  
Db 88 HLCDAQOHLQHGAFLTDYNNQADTTWQSQTMLAGVQYPNSINLTLLHLGKAFTYVVR 147  
QY 162 ----SRPDGC-----NNSSWIP-----RGNYESNR-----DD---CT--- 187  
Db 148 LKFTSRPESFAIYKRTREDGWPWPYQYSGSCENTYSKANRGFI RTGGDEQALCTDEF 207  
QY 188 --VSLIYAVHL-----KSGYVEFFEYQYVDNNIFFEFTIQDQCQEMDTTDDKWKVLT 238  
Db 208 SDISPLTGGNVAESTLEGRPSAYNF-----DN-----SPVLQEWVTATDIRVTLN 252  
QY 239 DNGEWG----SHSVMLKSGTNILYWRTTGILMG-----SKAVKPVLVKNI-TIEGV 284  
Db 253 RLNTFGDEVFNEPKVLKS---YYAISDFAVGGRCKCNGHASECVKNEFDKLMCNCKHN 308  
QY 285 AYTSECFPPCKPGTFSNKP-----GTFNCOVCPRNTYSEKGAKECTR 325  
Db 309 TYGVDCCKCLP-FFNDRPWRRRATAESASESLPCDCNGRSQECYFDELYRSTGHGGHCTN 367  
QY 326 CKDDSFSGSSECTER-----PPCTTKDYFOIHTPCDEEGKTQIMYKWIPEKICR 375  
Db 368 CRDNTDGAKCERCENFFRLGNTFACSPCHCSPVGLSTQCDSYGRCS-----CK 417  
QY 376 EDLTDATRLPPSGEKKDCPPCPNGFYNGSSSCHPCPPGTFSDGTKECR-----PCPAG 429  
Db 418 PGV-----MGDK--CDRCQPGHSLTEAGCRPC-SCDLRGSTDECNVETGRCVCKDN 466  
QY 430 TEPALGPEYK-----WNVLPGNMK--TSCFNVGNSK-CDGMNGWEV-----AG 470  
Db 467 VE---GFCERCKPGFFNLESSNPKGCTPCFCFHHSSVCTNAVGYSDISSTFOIDEDG 523  
QY 471 DHIOGAG-----GSDND--YLLNLHIPG-FKPPTSMTGATGSELGR-ITFVF---- 515  
Db 524 WRVEQRDSEASLEWSSDRQDIAVISDSYFFRYFIAPVKFLGNQVLSYGONLSFSERVDR 583  
QY 516 -ETLCSADCVLVFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW-----AF 567  
Db 584 RDTRLSAEDLVLEGAGL-RVSVPLIAQGNYSYPSETTVKYIFRLHEATDYPWRPALSPFEF 642  
QY 568 QRTNQGQDNRRFINDMVKI-----YSITATNAVIGVASSCRACALGSEQSGSCVPCPPG 622  
Db 643 Q-----KLLNLTSIKIRGYSERTAGYLDV-----TLQSAAPG-----PGVPA 682  
QY 623 HYIEKETNOCKECPDPTYLSIHQVYKFAKIPCGPGSKNNQDHSVCYSDCFFY----- 675  
Db 683 TWVESCT-----CPVG-----YGGQFCETCLPGYRRETPSLGPYSPCVLCTCNGHSE 729  
QY 676 -----HEKENQLLHYDFSNLSSVGLMNGPSETSKGTKYFHFFNISL 717  
Db 730 TCDPETGVCDNRDNTAGPHCEKCSGDIYGDSTLGTSSDCQPCPCPGSSSCAIVPKTEVV 789  
QY 718 CGH-----EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFVCCQSTIIPSESKGFRAAL 772  
Db 790 CTHCPTGTAGKRCELCDDGYEGDPL-----GSGNPVRLCRPCQCNDNDIDPN-AVGNCNRL 843  
QY 773 SSQSIILADTFIGVTVETTLKNINIKEDMEFPVPTSQIPDVHFFYKSSTATTSINGRSTA 832  
Db 844 TGECLKCIYNTAGFYCD-----RCKEGEFGNPLAPNP-----ADCKACACNPYGTIVQ 891  
QY 833 VKMRCNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESAECPLCTEHDHFEIEGAC---KR 890  
Db 892 QQSSCNPVGTQCQCLPHVSGRDCGTC--PGYVNLQSGQGCRCDCDCHALGSTNGQCQDIRT 949  
QY 891 GFQETLYVWNEPKWCIKGISLPEKKLATCETVDF 924  
Db 950 GQCE-----CQPGIT--GQHCERCETNHF 971

RESULT 15  
A57278  
fibrillin-2 precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 23-Feb-1996 #sequence\_revision 23-Feb-1996 #text\_change 02-Aug-2002  
C:Accession: A57278  
R:Zhang, H.; Hu, W.; Ramirez, F.  
J. Cell Biol. 129, 1165-1176, 1995  
A:Title: Developmental expression of fibrillin genes suggests heterogeneity of extracellular matrix  
A:Reference number: A57278; MUID:95263670; PMID:7744963  
A:Accession: A57278  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2907 <ZHA>  
A:Cross-references: GB:L39790; NID:g762830; PIDN:AAA74908.1; PID:g762831  
C:Superfamily: fibrillin 1; EGF homology  
F:1239-1274/Domain: EGF homology <EGF1>  
F:2488-2523/Domain: EGF homology <EGF>

Query Match 3.3%; Score 177.5; DB 2; Length 2907;  
Best Local Similarity 20.1%; Pred. No. 0.0031;  
Matches 224; Conservative 101; Mismatches 375; Indels 417; Gaps 67;

QY	41	GCQAAWAGDL-----PSSSRPLPPCOEKDYHFEYTECDSSGSRWR--VAIPNSAVDC	91
Db	17	GCVALWAQTDGQPPPPKTLWPQPPQVRPAVAGSEGGFMPEYRDEGAVRASRVR	76
QY	92	SGLPDPVRGKE-C-----TFSCASGEYLEMKNQ-----VC-SKCGEGTYS-----	129
Db	77	RQQEILRGNVCGSRFHSYCCPGWKTLPGGNQCIVIPICRNSCGDGFCSRPNMCTCSSGQ	136
QY	130	-----LGSQIKFDEWDELPAQFSNIATFMDTVVGPDSR-----DGCNN	169
Db	137	ISPTCGRKSIOQCSVRCMNGGTCAADHCOCQKGY-----IGTYCGQPVCEGCGQN	186
QY	170	SSWIPRGNYIESNRDDCTVSLIYAVHLKKSQYVF-----FEYQYVDNNIFFEFFFIONDQCQ	225
Db	187	G-----GRCIGPNRCACV-----YGFTGPQCDERYTGPCFTQ--VNNQMCQ	226
QY	226	EMDT---TTDKWKVKLTDNGEGWSHVMKSGTNILYWRRTTGILMGSKAVKPVLVKNITIE	282
Db	227	GOLTGIVCTKTLCCATIGRAWGHPCEMCPA-----QPQCRPGFIPNIRTG	272
QY	283	GVAYTSEC--FP--CKPGTFESNKGCSFNQVCPNRTYSEKGAKECIRCKDDSQFSGSSEC	338
Db	273	ACQDVDECAIPGLCQGGNCINTVGSFECP--AGHKQ-----SET	312
QY	339	TERPPCTTKDYFQIHTPCDEEGTQIMYKWIPIKICREDLTDAILRPPSGEKKDCPPCNP	398
Db	313	TQK-----C-EDIDECVPIPGVCEGTDC-----	334
QY	399	GFYNN--GSSSCHPCPPG--TFSDGTEKCRPCPAGTEPALGFYKWNVLPGNMKTS---	451
Db	335	---SNTVGSYFC-LCPRGEVSTDGSRCIDQAGTCFSGLVNGRCAQELPGRMAKAQCC	389
QY	452	-----CFNVGN--SKC-----DGMNGWEVAGDHIQ--SGAGGSDNDYLILN-	488
Db	390	CEPGRCWSIGTIPEACPVRGSEYRRLCLDGLPMGGIPGSSVSRRPGGTSGNNGNGYGPGGT	449
QY	489	--LHIP--GFKPPTSMTGATGSLGRI---TFVFET-----LC-----	519
Db	450	GFLPIPDNGFSPGVGGAGVGAGGQGPITGLTILNQITIDICKHHANLCLNGRCIPTVSS	509
QY	520	-----SADCVLYFMVDINRKSTNVVESWGGTKEKQAY---THIIFKNATFTF	563
Db	510	YRCRCNMGYKQDANGDC-----IDVDECTSNPCNSGDCVNTPGSYCKCHA-----	555
QY	564	TWAFQRTNQGQ-----DNRRFINDMVKIYSI-----TATNAVD-GV	598
Db	556	--GFQRTPTKQACIDIDEQIONGVLCNKRVCNSDGSFQICINAGFELTTDGKNCVDHDE	613
QY	599	ASSCRACALG---SEQSGSSCVPCPPGPHYIEKETNQC---KEC-PPDTYLSIHQV--YGK	649

Db	614	CTTTNMCLNGMCINEDGSFKCV-CKPGFILAPNGRYCTDVDECQTPGICMNGHCINNEGS	672
QY	650	EACIPCGPGSKNNQDHSVCY-----SDCFFYHEKENQILHYDFSLLSVGSLMNGPSF--	702
Db	673	FRC-DCPPGLAVGVDRVCVDTHMRSTCYGFIKKGVCRPFPFGAVTKSECCCAN-PDYGF	730
QY	703	-----TSKGTQYFH-----FFNISLCGHEGKKMAL-----CTNNITDFTVKEIVAGSDD	746
Db	731	GEPQPCPAKNSAEFHGLCSSLGIGITVDRDINECALDPDICANGICE-----	778
QY	747	YTNLVGAFCVQSTIIPSESKGFRAALSSQSIILADTFIGTVTVETTLKNINIKED-----	800
Db	779	--NLRGTYRCNC-----NSGYEPDASGRNCIDIDECL---VNRLCDNGLCRNTPGSYS	827
QY	801	-----MFPVPTSOIPDVHFFYKSSATTATTSINGRST---AVKMRGNPTK--SGAGVI	847
Db	828	CTCPPGYVLPTEETCEDV-----NECESNPCVNGACRNNLGSFHCESPGSKLSSTGLI	882
QY	848	SVPSKCPAGTC-----DGCTFYF-----LWESAEACPLCTEHDFHEIEG	886
Db	883	CIDSL--KGTCLWLNQDNRCVNINGATLKSCECCATLGGAWGS--PCERC-----ELDA	932
QY	887	ACKRGFQETLYVWNEPKWCICKISLPEKKLATCETVD	923
Db	933	ACPRGFAR-----IKGV-----TCEDVN	950

Search completed: May 12, 2003, 13:21:11  
Job time : 47.0673 secs





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 11:12:58 ; Search time 19.3568 Seconds  
(without alignments)  
2063.447 Million cell updates/sec

Title: US-10-073-333A-2  
Perfect score: 5357  
Sequence: 1 MLFRARGPVRGRGWGRPAEA.....TCYFWKKKQKKKTILNLFN 963

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	242	4.5	1877	1 PCK5_MOUSE	Q04592 mus musculus
2	240	4.5	1696	1 PCK5_BRACL	Q9nj15 branchiosto
3	223.5	4.2	687	1 VS41_GIALA	P92127 giardia lam
4	210	3.9	713	1 TSA4_GIALA	P21849 giardia lam
5	200	3.7	3718	1 LMA5_MOUSE	Q61001 mus musculus
6	194	3.6	5376	1 ZAN_MOUSE	O88799 mus musculus
7	182	3.4	667	1 TS11_GIALA	Q03185 giardia lam
8	178.5	3.3	3672	1 LML2_CAEEL	Q21313 caenorhabdi
9	177.5	3.3	1607	1 LMG1_MOUSE	P02468 mus musculus
10	177.5	3.3	2907	1 FBN2_MOUSE	Q61555 mus musculus
11	173.5	3.2	1581	1 LMG3_MOUSE	Q9r0b6 mus musculus
12	173.5	3.2	2871	1 FBN1_MOUSE	Q61554 mus musculus
13	172	3.2	1786	1 LMB1_MOUSE	P02469 mus musculus
14	170.5	3.2	2437	1 NTC1_BRARE	P46530 brachydanio
15	170	3.2	1557	1 LML1_CAEEL	Q18823 caenorhabdi
16	169	3.2	1639	1 LMG1_DROME	P15215 drosophila
17	168	3.1	1786	1 LMB1_HUMAN	P07942 homo sapien
18	167.5	3.1	1587	1 LMG3_HUMAN	Q9y6n6 homo sapien
19	166	3.1	2871	1 FBN1_HUMAN	P35555 homo sapien
20	164.5	3.1	1790	1 LMB1_DROME	P11046 drosophila
21	164	3.1	2871	1 FBN1_BOVIN	P98133 bos taurus
22	164	3.1	2871	1 FBN1_PIG	Q9tv36 sus scrofa
23	164	3.1	3084	1 LMA1_MOUSE	P19137 mus musculus
24	164	3.1	3110	1 LMA2_HUMAN	P24043 homo sapien
25	161	3.0	3712	1 LMA_DROME	Q00174 drosophila
26	160.5	3.0	1609	1 LMG1_HUMAN	P11047 homo sapien
27	160.5	3.0	2911	1 FBN2_HUMAN	P35556 homo sapien
28	158	2.9	3695	1 LMA5_HUMAN	O15230 homo sapien
29	156.5	2.9	3106	1 LMA2_MOUSE	Q60675 mus musculus
30	155.5	2.9	2003	1 NTC4_HUMAN	Q99466 homo sapien
31	153.5	2.9	1680	1 FUR2_DROME	P30432 drosophila
32	153.5	2.9	2491	1 MPRI_HUMAN	P11717 homo sapien
33	152	2.8	3707	1 PGBM_MOUSE	Q05793 mus musculus

34	151.5	2.8	2703	1 NOTC_DROME	P07207 drosophila
35	151	2.8	2470	1 NTC2_MOUSE	O35516 mus musculu
36	148	2.8	2482	1 VWF_PIG	Q28833 sus scrofa
37	148	2.8	2531	1 NTC1_MOUSE	Q01705 mus musculu
38	148	2.8	2704	1 G168_PARPR	P17053 paramesium
39	147	2.7	2321	1 NTC3_HUMAN	Q9um47 homo sapien
40	145	2.7	610	1 LEM2_HUMAN	P16581 homo sapien
41	144	2.7	1816	1 LMA4_HUMAN	Q16363 homo sapien
42	144	2.7	2318	1 NTC3_MOUSE	Q61982 mus musculu
43	143.5	2.7	2524	1 NOTC_XENLA	P21783 xenopus lae
44	143	2.7	1231	1 CFAH_HUMAN	P08603 homo sapien
45	142.5	2.7	1959	1 AGRI_RAT	P25304 rattus norv

ALIGNMENTS

RESULT 1  
PCK5\_MOUSE  
ID PCK5\_MOUSE STANDARD; PRT; 1877 AA.  
AC Q04592; Q62040;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)  
DE (Proprotein convertase PC5) (Subtilisin/kexin-like protease PC5)  
DE (Convertase PC5) (PC6) (Subtilisin-like proprotein convertase 6)  
DE (SPC6).  
GN PCSK5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PC5B).  
RC STRAIN=ICR; TISSUE=Intestine;  
RX MEDLINE=93327934; PubMed=8335106;  
RA Nakagawa T., Murakami K., Nakayama K.;  
RT "Identification of an isoform with an extremely large Cys-rich region  
of PC6, a Kex2-like processing endoprotease.";  
RL FEBS Lett. 327:165-171(1993).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM PC5A).  
RC TISSUE=Brain, and Intestine;  
RX MEDLINE=93224489; PubMed=8468318;  
RA Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K.,  
RA Nakayama K.;  
RT "Identification and functional expression of a new member of the  
mammalian Kex2-like processing endoprotease family: its striking  
structural similarity to PACE4.";  
RL J. Biochem. 113:132-135(1993).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM PC5A).  
RC TISSUE=Adrenal cortex;  
RX MEDLINE=93342056; PubMed=8341687;  
RA Lussan J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;  
RT "cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a  
candidate proprotein convertase expressed in endocrine and  
nonendocrine cells.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).  
RN [4]  
RP PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.  
RX MEDLINE=97103178; PubMed=8947550;  
RA De Bie I., Marcinkiewicz M., Malide D., Lazure C., Nakayama K.,  
RA Bendayan M., Seidah N.G.;  
RT "The isoforms of proprotein convertase PC5 are sorted to different  
subcellular compartments.";  
RL J. Cell Biol. 135:1261-1275(1996).  
RN [5]  
RP DEVELOPMENTAL EXPRESSION.  
RX MEDLINE=96293359; PubMed=8698813;  
RA Constam D.B., Calton M., Robertson E.J.;  
RT "SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone

RT morphogenetic proteins at distinct sites during embryogenesis.";

RL J. Cell Biol. 134:181-191(1996).

RN [6]

RP DEVELOPMENTAL EXPRESSION.

RX MEDLINE-97436919; PubMed-9291583;

RA Rancourt S.L., Rancourt D.E.;

RT "Murine subtilisin-like proteinase SPC6 is expressed during embryonic implantation, somitogenesis, and skeletal formation.";

RL Dev. Genet. 21:75-81(1997).

CC -!- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF. MAY BE RESPONSIBLE FOR THE MATURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION OF GROWTH FACTORS.

CC -!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR PROTEINOLYTI BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS, WHERE XAA CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.

CC -!- SUBCELLULAR LOCATION: PC5A IS SECRETED THROUGH THE REGULATED SECRETORY PATHWAY. PC5B IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH EARLY ENDOSOMES.

CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; PC5B/LONG (SHOWN HERE) AND PC5A/SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -!- TISSUE SPECIFICITY: PC5A IS EXPRESSED IN MOST TISSUES BUT IS MOST ABUNDANT IN THE INTESTINE AND ADRENALS. PC5B IS EXPRESSED IN THE INTESTINE, ADRENALS AND LUNG BUT NOT IN THE BRAIN.

CC -!- DEVELOPMENTAL STAGE: WEAKLY EXPRESSED THROUGHOUT THE EMBRYO, EXCEPT IN THE DEVELOPING NERVOUS SYSTEM, THE RIBS AND THE LIVER, BUT MARKEDLY UPREGULATED AT DISCRETE SITES DURING DEVELOPMENT. AT E6.5, PROMINENT EXPRESSION OBSERVED IN DIFFERENTIATED DECIDUA. AT E7.5, INTENSE EXPRESSION IN EXTRAEMBRYONIC ENDODERM, AMNION AND NASCENT MESODERM. AT 8.5, ABUNDANT EXPRESSION IN SOMITES AND YOLK SAC FOLLOWED BY A CONFINATION TO DERMATOTOME COMPARTMENT. BETWEEN E9.5 AND E11.5, ABUNDANT EXPRESSION IN AER (THICKENED ECTODERMAL CELLS OF LIMB BUDS). AT E12.5, EXPRESSION IN THE LIMBS IS CONFINED TO THE CONDENSING MESENCHYM SURROUNDING THE CARTILAGE. AT THIS STAGE, STRONG EXPRESSION ALSO DETECTED IN VERTEBRAL AND FACIAL CARTILAGE PRIMORDIA AND IN THE MUSCLE OF THE TONGUE. AT E16.5, ABUNDANT EXPRESSION IN EPITHELIAL CELLS OF THE INTESTINAL VILLI. ISOFORM A IS MOST ABUNDANT AT ALL STAGES BUT SIGNIFICANT LEVELS OF ISOFORM B OCCUR AT E12.5.

CC -!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE ASSISTING THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC RETICULUM.

CC -!- DOMAIN: AC 1 AND AC 2 (CLUSTERS OF ACIDIC AMINO ACIDS) CONTAIN SORTING INFORMATION. AC 1 DIRECTS TGN LOCALIZATION AND INTERACTS WITH THE TGN SORTING PROTEIN PACS-1.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.

CC -!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; D17583; BAA04507.1; -

CC EMBL; D12619; BAA02143.1; -

CC EMBL; L14932; AAA74636.1; -

CC PIR; JX0248; JX0248.

CC PIR; A48225; A48225.

CC HSSP; Q99405; 1MPT.

CC MEROPS; S08.076; -

CC MGD; MGI:97515; pcks5.

CC InterPro; IPR000561; EGF-like.

CC InterPro; IPR002174; Furin-like.

CC InterPro; IPR002884; P\_domain.

CC InterPro; IPR000209; Peptidase\_S8.

CC Pfam; PF00082; Peptidase\_S8; 1.

CC Pfam; PF01483; P; PARTIAL.

DR PRINTS; PR00723; SUBTILISIN.

DR ProDom; PD000717; P\_domain; 1.

DR SMART; SM00181; EGF; 3.

DR SMART; SM00001; EGF\_like; 2.

DR SMART; SM00261; FU; 22.

DR PROSITE; PS00136; SUBTILASE\_ASP; 1.

DR PROSITE; PS00137; SUBTILASE\_HIS; 1.

DR PROSITE; PS00138; SUBTILASE\_SER; 1.

KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;

KW Cleavage on pair of basic residues; Repeat; Alternative splicing;

KW Transmembrane.

FT SIGNAL 1 34

FT PROPEP 35 116

FT CHAIN 117 1877

FT

FT DOMAIN 117 1768

FT TRANSMEM 1769 1789

FT DOMAIN 1790 1877

FT DOMAIN 117 452

FT DOMAIN 464 602

FT DOMAIN 638 1753

FT DOMAIN 1825 1844

FT DOMAIN 1856 1877

FT SITE 116 117

FT SITE 521 523

FT ACT\_SITE 173 173

FT ACT\_SITE 214 214

FT ACT\_SITE 388 388

FT CARBOHYD 227 227

FT CARBOHYD 383 383

FT CARBOHYD 667 667

FT CARBOHYD 754 754

FT CARBOHYD 804 804

FT CARBOHYD 854 854

FT CARBOHYD 951 951

FT CARBOHYD 1016 1016

FT CARBOHYD 1220 1220

FT CARBOHYD 1317 1317

FT CARBOHYD 1523 1523

FT CARBOHYD 1711 1711

FT CARBOHYD 1733 1733

FT VARSPLIC 878 915

FT

FT VARSPLIC 916 1877

FT SEQUENCE 1877 AA; 209287 MW; EC850E2DF20EA1C3 CRC64;

SQ

Query Match 4.5%; Score 242; DB 1; Length 1877;

Best Local Similarity 19.1%; Pred. No. 1.2e-09;

Matches 221; Conservative 97; Mismatches 382; Indels 456; Gaps 64;

Qy 61 CQEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKEC-----TFSCA 108

Db 676 CPPGHYHADKKRC-----RKCAPN-CESCFG----SHGQCLSKYGYFLNEETSSCV 723

Qy 109 S----GEYLEMKNQVCSKCGEGYSLGSGIKFDEWDELPAFGS-----NIATFMDTVVG 158

Db 724 TQCPDGSYEDIKKNVCGKCSNCKAC---IGFHNCTECKGGLSLQGRCSVTCDCQFFN 780

Qy 159 PSDSRP-----DCNNSSWIPRGNIEYNR--DDCTVSLIYAVHLKKSQYVF 203

Db 781 GHDCQPCHRFCATCSGAGADGCINCT---EGYVMEGRGCVQSCSVS--YYLDHSSEGGYK- 835

Qy 204 FEQYVVDNNIFF-----EFTQNDQCEHMDTTF 231

Db 836 -SCKRCDNSCLTCNGPGFKNCSSCPSGYLLDLGTGCOMGAICKDGEYIDDGHCQTCEASC 894

Qy 232 DKWVKLTNGEWGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECF 291

Db 895 AKC-----WG-----PTQEDCISCVPVTRVLDGRC-----VMNCPSWKFEPKQCH 935

Qy 292 PCKPGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDD-----SQFSGSSECTERPPCTTK 347

```
Db 936 PC-----HYTCQGC-----QSGSPSNTSCRADKHGQERFLYHGECLN--CPVG 978
Qy 348 DY-FQIHT--PCDEEGKTQIMYKWIEPKICREDLTDAILRPPS-----GE 389
Db 979 HYPAGHGTCLPCPD--NCELCY---NPHICSRMSGYVIIPNHTCQKLECRQGEFQDSE 1033
Qy 390 KKDCCPCNPGFY---NNGSSSCHPCPPGTFSDGTKECRPCAGTEPALGFEYKWNVLPG 446
Db 1034 YEECMPCBEGCLGCTEDDPGACTSCATGYMFERHCYKACPEKT-----FGVKECRACG 1088
Qy 447 NMKTSCFNVGNSKCDGMNGWEVAGDHQSGAGGSDNDYLILNLHPIPGFKPPTSMGTATGS 506
Db 1089 ---TNGSCDQHEC---YWCCEGFFLGGSCVQDCG-----PGFH-----GDQ 1125
Qy 507 ELGRITTFETLCSADCVLYFMVDINPKST--NVVESWGGTKEKEQAYTHIIFKNATFTFT 564
Db 1126 ELGEC-----KPCRACETCTGSGYNQCSSCQEGQLQHLWGTCLWSTWPOVEGKD----- 1174
Qy 565 WAFORTNQ-----QDNRRFNDMVKIYSITATNAVDGVASSCRACALGSEQ 611
Db 1175 W-----NEAVPTEKPSLVRSLLDQRRKW--KVQI-KRDATSONQPCCHSSCKTC----- 1219
Qy 612 SGSSCVPCPPGHYI-----EKETNQCKEC-----PPD 638
Db 1220 NGSLCASCPTGMYLWLQACVPSCPQGTWPSVTSGSCEKCEDCVSCGADLCQQCLSQPD 1279
Qy 639 TYLSIHQ-----VYGKE-ACIPCGPGSKNNQ-----DHSVY--- 669
Db 1280 NTLLEHRCYHSCPEGFYAKDGVCEHCSSPCKTCEGNATSCNCEGDFVLDHGVCKWKC 1339
Qy 670 -----SDCFFYHEK-----ENQILHYDFSLSVGSGL----- 696
Db 1340 PEKHVAVEGVCKHCPERCQDCI--HEKTCKECMPDFFLYNDMCHSPCKSFYPDMRQCVP 1397
Qy 697 -----MNGPSFTSGTKYFHHFNISLCGHEGKKM--ALCTNNITDFTVKE----- 739
Db 1398 CHKNCLCNGPKED-----DCKVCADTSKALHNLGLDCECPGTYKEEENDECRD 1447
Qy 740 ----IVAGSDDYTNLV-----GAFVCQSTIIPSESKGFR 769
Db 1448 CPESCLICSSAWTCLACREGFTVVHDVCTAPKECAAVEYWDEGSHRCQ----PCHKKCSR 1503
Qy 770 AALSSQSIILADTFIGTVTETTLKNIN-IKEDMPFPVPTSQIPDVHFFYKSSATTSC--- 825
Db 1504 CSGPSE-----DQCYTCPRETFLNNTCVKE-----CPEGVHTDKDSQQCVLCHSS 1549
Qy 826 ---INGRSTAVKMRCNPTKSGAGVISVPSKC-----PAGTCDGC----- 861
Db 1550 CRTCEGPHSMQCLSCR-----GWFLGKECLLQCRDGYVYGESTSGRCEKCDKCKSCRG 1604
Qy 862 -----TFYFLWESAEC-PLCTEHDF-----HEIEGAC--KRGFQETLY 897
Db 1605 PRPTDCQSCDTFFFLRLRSKGQCHRACPEHYADQHAQTCCERCHPTCDKCSGKEAWSCLSC 1664
Qy 898 VVNEPKWCIKGISLPE 913
Db 1665 VWS--YHLKIGICIPE 1678

RESULT 2
PCK5_BRACL
ID PCK5_BRACL STANDARD; PRT; 1696 AA.
AC Q9NJ15; Q9NJ16; Q9NJ14;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
DE (Proprotein convertase PC6-like) (aPC6).
GN PC6.
OS Branchiostoma californiensis (California lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7738;
```

```

[1]
SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
MEDLINE=20175281; Pubmed=10708868;
Oliva A.A. Jr., Chan S.J., Steiner D.F.;
"Evolution of the prohormone convertases: identification of a
homologue of PC6 in the protochordate amphioxus.";
Biochim. Biophys. Acta 1477:338-348(2000).
-!- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY
WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF (BY SIMILARITY).
-!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
PROPROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS, WHERE XAA
CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
-!- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.
ISOFORM B IS A TYPE I MEMBRANE PROTEIN.
-!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A, B (SHOWN HERE) AND C; ARE
PRODUCED BY ALTERNATIVE SPLICING.
-!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
ASSISTING THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC
RETICULUM.
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
-!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF184615; AAF26300.1; -
EMBL; AF184616; AAF26301.1; -
EMBL; AF184617; AAF26302.1; -
HSSP; Q99405; 1MPT.
MEROPS; S08.UPB; -.
InterPro; IPR000561; EGF-like.
InterPro; IPR002174; Furin-like.
InterPro; IPR002884; P_domain.
InterPro; IPR000209; Peptidase_S8.
Pfam; PF00082; Peptidase_S8; 1.
Pfam; PF01483; P; 1.
PRINTS; PR00723; SUBTILISIN.
ProDom; PD000717; P_domain; 1.
SMART; SM00181; EGF; 2.
SMART; SM00261; FU; 17.
PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
PROSITE; PS00137; SUBTILASE_HIS; 1.
PROSITE; PS00138; SUBTILASE_SER; 1.
Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
Cleavage on pair of basic residues; Repeat; Alternative splicing;
Transmembrane.
SIGNAL 1 25 POTENTIAL.
PROPEP 26 110 POTENTIAL.
CHAIN 111 1696 PROPROTEIN CONVERTASE SUBTILISIN/KEXIN
TYPE 5.
DOMAIN 111 1618 EXTRACELLULAR (POTENTIAL).
TRANSMEM 1619 1639 POTENTIAL.
DOMAIN 1640 1696 CYTOPLASMIC (POTENTIAL).
DOMAIN 111 488 CATALYTIC.
DOMAIN 496 637 HOMO B.
DOMAIN 664 1649 CYS-RICH MOTIF (CRM) REGION.
SITE 110 111 CLEAVAGE (AUTO-) (BY SIMILARITY).
ACT_SITE 192 192 CHARGE RELAY SYSTEM (BY SIMILARITY).
ACT_SITE 233 233 CHARGE RELAY SYSTEM (BY SIMILARITY).
ACT_SITE 407 407 CHARGE RELAY SYSTEM (BY SIMILARITY).
CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 885 885 N-LINKED (GLCNAC. . .) (POTENTIAL).
VARSPLIC 1259 1323 DDTLDRGECITSCGPGEYMDRREKKKACHPTCKECSDEY
DDTCTACNDGFLLTDDASCEAGCP -> AENQNSFCPFA
PREVSVLAELALGHLRLYSLTDPQSNPPDVTGLGADRRL
TTATSAAGRCA (IN ISOFORM C).
```



FT	VARSPLIC	1324	1696	MISSING (IN ISOFORM C).	
FT	VARSPLIC	1288	1343	CHPTCKECSDEYDDTCTACNDGFLLTLDASSCEAGCPPGQFL	
FT				HHGDCSCHRECKTC -> IARCVDRRDRSWCDLVLRFNFC	
FT				VRRYFVKRCCGCTCKLYMEDRPMRRGSSQPTQGRN (IN ISOFORM A).	
FT	VARSPLIC	1344	1696	MISSING (IN ISOFORM A).	
SQ	SEQUENCE	1696	AA; 188410	MW; 281CBE1784257CBD CRC64;	
Query Match 4.5%; Score 240; DB 1; Length 1696;					
Best Local Similarity 19.1%; Pred. No. 1.5e-09;					
Matches 206; Conservative 98; Mismatches 358; Indels 414; Gaps 57;					
QY	35	CCWALAGCAAWAGDLPSSSSRPLPCQEKDYHFE-----YTECDSSGSRWRVAIPNSAV	89		
DB	680	CHDSCATCHGRHSGQLS-----CHEGNYFEVEDEGTCSEECGQ--GYKDEEEKCL	729		
QY	90	DCSG--LPDPVRGKECTFSCASGEYLEMKNQVC-SKCGEGTYSLGSGIKFDEWDELPA	146		
DB	730	DCSADCLTCQVSADHCT-SCDDEGLKLFENTCVAQCSEGRY-----MDENDVC----	777		
QY	147	SNIAFEMDTVVGPSDSRPDGCNNSWIPRGNVIESNRDDCTVSLIYAVHLKKS	206		
DB	778	QDCDDSCDTCTGPDATDCVTCADEDLLQESQCVES-----CS-----SGYFQOEY	822		
QY	207	QYVDNNIFFEFF--IQNDQC-----QEMDTTDDKWKLTNDNGWGS	251		
DB	823	ECLKCHATCASCSGSRDDQCLTCSGHLELDEDTHRCITSCDEGEYGEKGKCEDC	882		
QY	252	S-----GTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECFFCKPGT	297		
DB	883	KCGSQADQCLECHHDN-LYDITTCVOYCGNRRYPE-----NGECHPCHPSC	928		
QY	298	FSNKPGSFN-----CQV-CPRNTYSEK-----AKECIRCKDD	329		
DB	929	LGCIGGEINQCNQCITDYEGEDHFLYQGTCHVTCPPLGYGDTTDQVCKACAPGCI	988		
QY	330	SQFSGSSECT---ERPCTTKDYFQIHTPCDEEGKTQIMYKWIE-----PKI---C	374		
DB	989	A---DNQCTLCEERAPT DGR-----CQSEGSQTDEAECAEGCHSCEGPDICD	1036		
QY	375	RED--LTD--AIRLP-----PSGEKKDCPPCN-----PGFYNNGS	405		
DB	1037	DEDYLLTDCVRRTRNCPSFTYPPDDQDRECRPCHDNCEACDGNPNQNCNSCKE	1096		
QY	406	SSCHPCPGTFSFGT-KECRPCPAGTEPALGFYKWNVLPGNMKTS CFNVGNSKCDGM	464		
DB	1097	GCSTGCPNRYKDDTNKECKPC-----DSSCFT-----CSG--	1127		
QY	465	GWEVAGDHQSGAGGSDNDYLILNLHHPGFKPPTSMTGATGSELGRITFVFETLCS	524		
DB	1128	---PASFHCLSCADGD-----FLHSSCRSTCP	1152		
QY	525	LYFMVDINRKSTNVVESWGGTKQKQAYTHIIFKNATFTFTWAFORTNQGQ-----	575		
DB	1153	AGFIG--NAFSEHCVES--SCEQDQYYS-----SETGRCEDCPYNCRACD	1193		
QY	576	NRFFINDMVKIYSI-----TATNAVDGVASSCRACALGSEQSG	613		
DB	1194	NGDCAECAPTYYIVDGRPCRPEETCEDGEYQDRDRDTAELSCRCHQSKCTCSG	1253		
QY	614	SSC-----VPCPPGHYIEKETNQCKECP-----DTYLSIHQVYGKEACIP	654		
DB	1254	DSCKGDDTILDRGECITSCGPGGEYMDRREKKCKACHPTCKECSDEY-----	1305		
QY	655	CGPGSKNQDHSVCYSDCFFYHEKENQILHY-DFSNLSSVGSMLNMGPSFTSKG	713		
DB	1306	CNDGFLLT-DASSCEAGC-----PPGQFLHHGDCDCHRECKTCDGP-----	1349		
QY	714	NISLCGHEGKKM--ALCTNNITDFTVKEIVAGSDDYTNLVGAFVCQSTIIPSE	771		
DB	1350	NCLSC-QPGSYLNDQQCSTHCPEGTPEE-----TYEDDSGETVLQCLCHVNCK	1397		
QY	772	LSSQSIILADTFIGTVETTLK---NINIKEDMFPVPTSQIPDVHFFYKSSSTAT	828		

DB	1398	-----TCHGEEDCMECANDIKYKQD-----GRCVTECQEG	1429		
QY	829	RSTAVKMRGNPTKSGAGVISVPSKCPACTGCTGFYFLWESAECAPLCTEHDHFEI	887		
DB	1430	HYPDLTNECQCQWS-----DCETCDG-----PRNDQCVC-PYNYLVLGKC	1470		
QY	888	---CKRGFQETLYVWNEPKWCIKGISLPEKKLATCE-----TVDFWLKV	932		
DB	1471	LEDCEGYDTMRQKECGEC-----HPSCATCNEGGNYNCLSCPYGSKLGEV	1519		
RESULT 3					
VS41_GIALA					
ID	VS41_GIALA	STANDARD;	PRT;	687	AA.
AC	P92127;				
DT	16-OCT-2001	(Rel. 40, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Variant-specific surface protein VSP4A1 precursor (CRISP-90).				
OS	Giardia lamblia (Giardia intestinalis).				
OC	Eukaryota; Diplomonadida; Hexamitidae; Giardia.				
OX	NCBI_TaxID=5741;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=O2-4A1;				
RX	MEDLINE=97321554; PubMed=9178264;				
RA	Papanastasiou P., Bruderer T., Li Y., Bommeli C., Koehler P.;				
RT	"Primary structure and biochemical properties of a variant-specific surface protein of Giardia."				
RL	Mol. Biochem. Parasitol. 86:13-27(1997).				
RN	[2]				
RP	CHARACTERIZATION.				
RX	MEDLINE=97233006; PubMed=9078242;				
RA	Papanastasiou P., McConville M.J., Ralton J., Koehler P.;				
RT	"The variant-specific surface protein of Giardia, VSP4A1, is a glycosylated and palmitoylated protein."				
RL	Biochem. J. 322:49-56(1997).				
CC	-!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE PLASMA MEMBRANE.				
CC	-!- PTM: O-GLYCOSYLATED. THE MAJOR GLYCAN IS A TRISACCHARIDE WITH GLC AT THE REDUCING TERMINUS.				
CC	-!- PTM: PALMITOYLATED.				
CC	-!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL; 283743; CAB06038.1; -.				
DR	HSSP; O14763; IDOG.				
DR	GlycoSuiteDB; P92127; -.				
DR	InterPro; IPR000561; EGF-like.				
DR	InterPro; IPR002174; Furin-like.				
DR	InterPro; IPR005127; Giardia_vsp.				
DR	Pfam; PF03302; VSP; 2.				
DR	SMART; SM00181; EGF; 2.				
DR	SMART; SM00001; EGF_like; 1.				
DR	SMART; SM00261; FU; 3.				
KW	Antigen; Repeat; Transmembrane; Glycoprotein; Lipoprotein; Palmitate; Signal.				
KW	Signal.				
FT	SIGNAL	1	14	POTENTIAL.	
FT	CHAIN	15	687	VARIANT-SPECIFIC SURFACE PROTEIN VSP4A1.	
FT	DOMAIN	15	660	EXTRACELLULAR (POTENTIAL).	
FT	TRANSMEM	661	681	POTENTIAL.	
FT	DOMAIN	682	687	CYTOSOLASMIC (POTENTIAL).	
SQ	SEQUENCE	687	AA; 70857	MW; D892F675D626D7EC	CRC64;
Query Match				4.2%;	Score 223.5; DB 1; Length 687;



Best Local Similarity 18.7%; Pred. No. 7.9e-09;  
Matches 144; Conservative 88; Mismatches 255; Indels 283; Gaps 36;

QY 290 CFPCKPGTF-----SNKPGSFNCQVCPRTYSEKAKECIRCKDDSQFSGSSECTE 340  
Db 83 CGKCGDGYFLFMGCGYKTESQPGS---EIC---TTASNGL--CTACKVDSQY----- 126

QY 341 RPPCTTKDYFQIHPCDEEGKTOIMYKWIPIKICREDLTDAILRPSGEKKKDCPPCNPGE 400  
Db 127 -----IFQKATPSEKSECILC-W-----DSTD--RNGVMG-VANCACTATA 166

QY 401 YNNGSSSCHPCPPGTFSDGTKECRPCPCAGTEPALGFYKWNVLPNGNMKTSFCNVGNSKC 460  
Db 167 SSTGPATCTECMAGTYKKSDTECAACHSDC-----ATCSGEANNQC 207

QY 461 DGMNGWEVAGHIQSGAGGSDNDYLILNLHIPGFKPPTSMT--GATGSELGRITVFET- 517  
Db 208 TSCE----TGKYLKSNQCVEKN-----TCNTNHYPDSTMTVCVACTVLDANCATCSFDSA 258

QY 518 -----LCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTWAFQT 570  
Db 259 TAKGKCLTCNSKNIPRTLD--GTSTCVENSYAGC----- 291

QY 571 NOQDNRFRIND---MVKIYSITATNAVGVASSCRACALGSEQSSSCVPCPPGHYIEK 627  
Db 292 -QGADNELFMKEDQSACLLCGDTKEASNDKGVANCRTCTKNANDSPPTCTACLDGYFLER 350

QY 628 ET-----NQCKECPDPTYLSIHQVYGKEACIPC----- 655  
Db 351 GSCTTTCADNATCSEATTEDKCKICKAGFFLASP---GEGKCISSDSTNNGGIDGCAEC 407

QY 656 -----GP-----GSKNN-----QDHSVYSCDFYHEKENQILHYDFSN 689  
Db 408 TKEPAGPLKCTKCKPNRPAGTSDNYTCTEDPTVC----- 446

QY 690 LSSVGLMNGPSTFTSGTKYFHFNFISLCGHEGKKM--ALCTNN-----ITDFTVKEIV 741  
Db 447 -GGTSGACDAIVIDANGKEHY--CSYCGETNKFPIDGLCTDNKGTNAGCTDHTCSYCA 502

QY 742 AGSDDY-----TNLVGAFVCQSTI-----IPSESKGF-----RAALSSQSIILADTFI 784  
Db 503 AGFFLYMGGYKIDTVPGSYMCSKSTAGVCDTPNANNRFEVVPKAIQSVLACGNPL 562

QY 785 GVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSSTAT-TSCINGRSTAVKMRNPTKSG 843  
Db 563 G-----TIAGGNAYVGVGCSQCTAPADRADGGMVATCTACEDGK-----KPGKSG 609

QY 844 AGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEHDFHEIEGACKRGFQETLYVWNEPK 903  
Db 610 TGCVA----CPDANCKSCT-----MDDVCEE----- 631

QY 904 WCIKGISLPEKKLATCETVDFWLKVAGVGAFATVL-----LVALTCYFW 948  
Db 632 -CADGFSLDNGKCVSSGNTKSGLSLGAIAGISVAIVVVGGLVGLCWWF 680

RESULT 4  
TSA4\_GIALA  
ID TSA4\_GIALA STANDARD; PRT; 713 AA.  
AC P21849;  
DT 01-MAY-1991 (Rel. 18, Created)  
DT 01-MAY-1991 (Rel. 18, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Major surface-labeled trophozoite antigen 417 precursor.  
GN TSA 417.  
OS Giardia lamblia (Giardia intestinalis).  
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.  
OX NCBI\_TaxID=5741;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 30957 / WB;  
RX MEDLINE=90280395; PubMed=2352929;  
RA Gillin F.D., Hagblom P., Harwood J., Aley S.B., Reiner D.S.,

McCaffery M., So M., Guiney D.G.;  
"Isolation and expression of the gene for a major surface protein of  
Giardia lamblia.";  
Proc. Natl. Acad. Sci. U.S.A. 87:4463-4467(1990).  
[2]  
SEQUENCE OF 480-620 FROM N.A.  
STRAIN=AD-1;  
MEDLINE=93314970; PubMed=8325510;  
Ey P.L., Mayrhofer G.;  
"Two genes encoding homologous 70-kDa surface proteins are present  
within individual trophozoites of the binucleate protozoan parasite  
Giardia intestinalis.";  
Gene 129:257-262(1993).  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE  
PLASMA MEMBRANE.  
CC -!- DOMAIN: CONTAINS 29 REPEATS OF THE CXXC MOTIF.  
CC -!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M33641; AAA02688.1; -;  
DR EMBL; M97488; AAA02581.1; -;  
DR PIR; A35502; A35502.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR005127; Giardia\_VSP.  
DR Pfam; PF03302; VSP; 2.  
DR SMART; SM00181; EGF; 1.  
DR SMART; SM00001; EGF\_like; 1.  
DR SMART; SM00261; FU; 3.  
KW Signal; Antigen; Glycoprotein; Transmembrane; Repeat.  
FT SIGNAL 1 17  
FT CHAIN 18 713 MAJOR SURFACE-LABELED TROPHOZOITE  
FT ANTIGEN 417.  
FT EXTRACELLULAR (POTENTIAL).  
FT POTENTIAL.  
FT CYTOPLASMIC (POTENTIAL).  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT A -> T (IN STRAIN ADELAIDE-1).  
FT VARIANT 582 606  
FT VARIANT 606 606  
FT SEQUENCE 713 AA; 72510 MW; 9AD7195843DE5601 CRC64;  
SQ

Query Match 3.9%; Score 210; DB 1; Length 713;  
Best Local Similarity 19.7%; Pred. No. 8.2e-08;  
Matches 186; Conservative 96; Mismatches 300; Indels 360; Gaps 53;

QY 90 DCSGLP-DPVRGKECTFSCASGEYLEMKNQ----VCSKCG-----EGTSLGSGIK-- 135  
Db 43 ECNGANYAPVNG-QCVDVNAEGPSKTLCPQHSAGKCTQCGGNSFMKYKDGCSYSSGGLPGH 101

QY 136 ----FDEWDEL-----PAGFSN-----IATFMDTVVGPST----S 162  
Db 102 SLCLSSDGDGVCTEAPGYFAPVGAANTEQSVIACGDTTGTVIAAGGNTYKGIADCAECS 161

QY 163 RPDGCNNSWIPRGNVIESNRDDCTVSLIYAVHLKKSQYVFFEQYVDNNIFFEFTQND 222  
Db 162 APDA---TAGAEAGKVATCTK--CGVS-----KYLKDNVCV-----D 193

QY 223 QCQEMDTTDDKWKVLTNDNGEWGSHSVMLKSGTNILYWRTTGILMGSKAVKPVLVKNITIE 282  
Db 194 KAQCNSGSTNKFVAV-DDSENGNKCVCSDNLN-----G 226

QY 283 GVAYTSECFPCPKPGTFSNKPGRFSCFNQVCPRTYSEKGAKECIRCKDDSQFSGSSE---C 338  
Db 227 GVA-----NCDTC---SYDEQSKKIKCTKCTDNNYLKTTSETGTC 263

Qy 339 TERPPCTTKDYFQIHPTPCDEEGTKQIMYKWIEPKICREDLTDAIRLPPSGEKKDCPPCNP 398  
Db 264 VQKQDC--KDGf---FPKDDSS-----AGNK--CLPCND 290  
Qy 399 GFYNNSSSCHPCP--PGTFSDGFKRCPCPAGTEPALGFYKWNVLPGNMKTSCFNVG 456  
Db 291 S--TDGIANCATCALVSGRGAALVTCsACTDGYKPSAD-----KTTCEAVS 335  
Qy 457 NSKCDGMNGWEVAGDHIQSGAGSDNDYLILNLHIPGFKPPTMTGATGSELGRITVFVE 516  
Db 336 NCKTPGCKACsNEGKENEVCtDCDSTYL-----TPISQ----- 369  
Qy 517 TLCSADCVLYFMVDINRKSTNVVSWGGTKEKQAYTHIIFKNATFTFTWAFQRT--NQGQ 574  
Db 370 --CIDSCA-----KIGNYyGATEGAKK-----LCRECTA----ANCKTCDDQGQ 407  
Qy 575 DNRFRINDMVKIYSITATNAVdGVASSCRACALGSEQSGSSVCPCPPGHYIEK-----ET 629  
Db 408 -----CQACNDGFYKNGDACSPC---HESCKTCSAGTA 437  
Qy 630 NOCKECPDPTYLSIHQVYGEACI-PCGPGSKNNQDNHsVCysDCFFYHEKENQILHYDFS 688  
Db 438 SDCTECPtGKALR---YGDdGtKGtCGEGCTTGTGAGAC----- 473  
Qy 689 NLSSVGSMLNGSPFTSKGtKYfHFFNLSLCGHEGKKMALCTNNITDFTVKEIVAG--SDD 746  
Db 474 --KTCGLTIDGAsYcSECATTTEYPONGVC---APKASRATPTCNDSPiQNGVCGTCADN 528  
Qy 747 YTNLVGAfVcQSTIIPSESKGFRAALSSQSIIADTFIGTVtETLKNINIKEDMFPVPT 806  
Db 529 YfKMNGG--CYETV--KYPGKTVCtSAPN-----GGTCQKAADGYKLDsGLTLVCS 575  
Qy 807 SQIPDVHfFYKSSTATtSCING--RSTAVKMRCN---PTKSGAGVISVPsKCPAGTCDCG 861  
Db 576 EGCKEC-----ASSTDCtTCLDGYVKsASACTKCDASCETCNGA-----ATTCKAC 621  
Qy 862 -TFYfLWESAE-ACPLCTEHDFHEIEGACKRGfQETLYVWNPEKWCiKGISLPEKK-LAT 918  
Db 622 ATGYyKTASGEgACTsC-ESDSNGVTGI-----KGCLNCAPPPNKGsVYL 665  
Qy 919 C-----ETVDFWLKVGAvgAFTAVL-----LVALTCYfW 948  
Db 666 CYLIKDSGSTNKsGLSTGAiAGISVAVIVVVGGLIGfLCWfW 707

RESULT 5

LMA5\_MOUSE  
ID LMA5\_MOUSE STANDARD; PRT; 3718 AA.  
AC Q61001; Q9JH06;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Laminin alpha-5 chain precursor.  
GN LAMA5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE OF 1-92 FROM N.A.  
RA Timpl R., Sasaki T.;  
RT "Completion of the N-terminal sequence of the murine Laminin alpha 5 chain.";  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 84-3718 FROM N.A.  
RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;  
RX MEDLINE=96081906; PubMed=7499364;  
RA Miner J.H., Lewis R.M., Sanes J.R.;  
RT "Molecular cloning of a novel laminin chain, alpha 5, and widespread expression in adult mouse tissues.";  
RL J. Biol. Chem. 270:28523-28526(1995).  
RN [3]

RP REVISIONS.  
RA Miner J.H., Lewis R.M., Sanes J.R.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin is thought to mediate the attachment, migration, and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components.  
CC -!- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.  
CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three different polypeptide chains (alpha, beta, gamma), which are bound to each other by disulfide bonds into a cross-shaped molecule comprising one long and three short arms with globules at each end.  
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT MEMBRANES (MAJOR COMPONENT).  
CC -!- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT AND SKIN. EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.  
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.  
CC -!- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.  
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
CC -!- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.  
CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AJ293593; CAB99255.1; -.  
DR EMBL; U37501; AAC53430.1; -.  
DR HSSP; P02468; 1TLE.  
DR MGD; MGI:105382; Lama5.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001886; LamNT.  
DR InterPro; IPR000034; Laminin\_B.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR InterPro; IPR001791; Laminin\_G.  
DR Pfam; PF00052; laminin\_B; 1.  
DR Pfam; PF00053; laminin\_EGF; 19.  
DR Pfam; PF00054; laminin\_G; 2.  
DR Pfam; PF00055; laminin\_Nterm; 1.  
DR ProDom; PD002082; LamNT; 1.  
DR ProDom; PD003031; Laminin\_B; 1.  
DR PROSITE; PS00022; EGF\_1; 19.  
DR PROSITE; PS01186; EGF\_2; 3.  
DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 19.  
DR PROSITE; PS00025; LAM\_G\_DOMAIN; 5.  
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.  
FT SIGNAL. 1 40 POTENTIAL.  
FT CHAIN 41 3718 LAMININ ALPHA-5 CHAIN.  
FT DOMAIN 41 304 LAMININ N-TERMINAL (DOMAIN VI).  
FT DOMAIN 305 363 LAMININ EGF-LIKE 1.  
FT DOMAIN 364 433 LAMININ EGF-LIKE 2.  
FT DOMAIN 434 479 LAMININ EGF-LIKE 3.  
FT DOMAIN 500 546 LAMININ EGF-LIKE 4.  
FT DOMAIN 547 592 LAMININ EGF-LIKE 5.  
FT DOMAIN 593 637 LAMININ EGF-LIKE 6.  
FT DOMAIN 638 682 LAMININ EGF-LIKE 7.  
FT DOMAIN 683 728 LAMININ EGF-LIKE 8.  
FT DOMAIN 729 781 LAMININ EGF-LIKE 9.  
FT DOMAIN 782 833 LAMININ EGF-LIKE 10.  
FT DOMAIN 834 855 LAMININ EGF-LIKE 11 (INCOMPLETE).  
FT DOMAIN 856 1442 LAMININ DOMAIN IV 1 (DOMAIN IV B).  
FT DOMAIN 1443 1488 LAMININ EGF-LIKE 12.  
FT DOMAIN 1489 1532 LAMININ EGF-LIKE 13.

FT DOMAIN 1533 LAMININ. EGF-LIKE 14.  
FT DOMAIN 1582 LAMININ EGF-LIKE 15.  
FT DOMAIN 1632 LAMININ EGF-LIKE 16 (N-TERMINAL).  
FT DOMAIN 1633 LAMININ DOMAIN IV 2 (DOMAIN IV A).  
FT DOMAIN 1643 LAMININ EGF-LIKE 17.  
FT DOMAIN 1832 LAMININ EGF-LIKE 16 (C-TERMINAL).  
FT DOMAIN 1864 LAMININ EGF-LIKE 17.  
FT DOMAIN 1865 LAMININ EGF-LIKE 18.  
FT DOMAIN 1914 LAMININ EGF-LIKE 19.  
FT DOMAIN 1970 LAMININ EGF-LIKE 20.  
FT DOMAIN 1971 LAMININ EGF-LIKE 21.  
FT DOMAIN 2024 LAMININ EGF-LIKE 22.  
FT DOMAIN 2071 DOMAIN II AND I.  
FT DOMAIN 2072 LAMININ G-LIKE 1.  
FT DOMAIN 2118 LAMININ G-LIKE 2.  
FT DOMAIN 2168 LAMININ G-LIKE 3.  
FT DOMAIN 2735 LAMININ G-LIKE 4.  
FT DOMAIN 2736 LAMININ G-LIKE 5.  
FT DOMAIN 2947 COILED COIL (POTENTIAL).  
FT DOMAIN 3128 COILED COIL (POTENTIAL).  
FT DOMAIN 3337 COILED COIL (POTENTIAL).  
FT DOMAIN 3518 COILED COIL (POTENTIAL).  
FT DOMAIN 2205 CELL ATTACHMENT SITE (POTENTIAL).  
FT DOMAIN 2330 CELL ATTACHMENT SITE (POTENTIAL).  
FT DOMAIN 2604 BY SIMILARITY.  
FT DOMAIN 2639 BY SIMILARITY.  
FT SITE 1723 BY SIMILARITY.  
FT SITE 1839 BY SIMILARITY.  
FT DISULFID 305 BY SIMILARITY.  
FT DISULFID 307 BY SIMILARITY.  
FT DISULFID 329 BY SIMILARITY.  
FT DISULFID 341 BY SIMILARITY.  
FT DISULFID 364 BY SIMILARITY.  
FT DISULFID 366 BY SIMILARITY.  
FT DISULFID 401 BY SIMILARITY.  
FT DISULFID 413 BY SIMILARITY.  
FT DISULFID 434 BY SIMILARITY.  
FT DISULFID 436 BY SIMILARITY.  
FT DISULFID 454 BY SIMILARITY.  
FT DISULFID 466 BY SIMILARITY.  
FT DISULFID 500 BY SIMILARITY.  
FT DISULFID 502 BY SIMILARITY.  
FT DISULFID 523 BY SIMILARITY.  
FT DISULFID 535 BY SIMILARITY.  
FT DISULFID 547 BY SIMILARITY.  
FT DISULFID 549 BY SIMILARITY.  
FT DISULFID 568 BY SIMILARITY.  
FT DISULFID 580 BY SIMILARITY.  
FT DISULFID 593 BY SIMILARITY.  
FT DISULFID 595 BY SIMILARITY.  
FT DISULFID 613 BY SIMILARITY.  
FT DISULFID 625 BY SIMILARITY.  
FT DISULFID 638 BY SIMILARITY.  
FT DISULFID 640 BY SIMILARITY.  
FT DISULFID 658 BY SIMILARITY.  
FT DISULFID 670 BY SIMILARITY.  
FT DISULFID 683 BY SIMILARITY.  
FT DISULFID 685 BY SIMILARITY.  
FT DISULFID 704 BY SIMILARITY.  
FT DISULFID 716 BY SIMILARITY.  
FT DISULFID 1443 BY SIMILARITY.  
FT DISULFID 1445 BY SIMILARITY.  
FT DISULFID 1464 BY SIMILARITY.  
FT DISULFID 1476 BY SIMILARITY.  
FT DISULFID 1533 BY SIMILARITY.  
FT DISULFID 1535 BY SIMILARITY.  
FT DISULFID 1557 BY SIMILARITY.  
FT DISULFID 1569 BY SIMILARITY.  
FT DISULFID 1582 BY SIMILARITY.  
FT DISULFID 1584 BY SIMILARITY.  
FT DISULFID 1603 BY SIMILARITY.  
FT DISULFID 1615 BY SIMILARITY.  
FT DISULFID 1630 BY SIMILARITY.  
FT DISULFID 1665 BY SIMILARITY.  
FT DISULFID 1867 BY SIMILARITY.  
FT DISULFID 1884 BY SIMILARITY.  
FT DISULFID 1896 BY SIMILARITY.  
FT DISULFID 1915 BY SIMILARITY.  
FT DISULFID 1917 BY SIMILARITY.

FT DISULFID 1941 1950 BY SIMILARITY.  
FT DISULFID 1953 1968 BY SIMILARITY.  
FT DISULFID 1971 1986 BY SIMILARITY.  
FT DISULFID 1973 1993 BY SIMILARITY.  
FT DISULFID 1996 2005 BY SIMILARITY.  
FT DISULFID 2008 2022 BY SIMILARITY.  
FT DISULFID 2072 2083 BY SIMILARITY.  
FT DISULFID 2074 2090 BY SIMILARITY.  
FT DISULFID 2092 2101 BY SIMILARITY.  
FT DISULFID 2104 2116 BY SIMILARITY.  
FT DISULFID 2119 2126 BY SIMILARITY.  
FT DISULFID 2121 2133 BY SIMILARITY.  
FT DISULFID 2135 2144 BY SIMILARITY.  
FT DISULFID 2147 2166 BY SIMILARITY.  
FT DISULFID 2169 2169 INTERCHAIN (PROBABLE).  
FT DISULFID 2172 2172 INTERCHAIN (PROBABLE).  
FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 905 905 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 926 926 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1335 1335 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2021 2021 N-LINKED (GLCNAC. . .) (POTENTIAL).  
  
Query Match 3.7%; Score 200; DB 1; Length 3718;  
Best Local Similarity 18.5%; Pred. No. 3.7e-06;  
Matches 174; Conservative 82; Mismatches 279; Indels 404; Gaps 46;  
  
QY 6 RGPVVRGRG-----WGRPAEAPRRGRSPWPSPAWICCWALAGCOAAWAGDLPSS---S 54  
Db 1465 RGHVIGRDCSRCATGYWGFNCRP-----CDGARGLCDELGTQCICP 1506  
  
QY 55 SRPLPP----CQEKDYHFE---YTECDSSG-SRWRVAIPNSAVD---CSGLPDPVRGKE 102  
Db 1507 PRTVPPDCLVCQPQSFQFGCHPLVGCCECNCSGPGVQELTDTCTCDMDSGQCRCPN-VAGRR 1565  
  
QY 103 CTFSCASGEY-----LEMKNOV-----CSKCGEGTYSLGS 132  
Db 1566 CD-TCAPGFYGYPCRPCDCHEAGTMASVCDPLTGQCHCKENVQSGRCDQCRVGTFSLDA 1624  
  
QY 133 G-----IKFDEWDELPAFNSIATF-----MDTVVGPSDSRP----- 164  
Db 1625 ANPKGCTRCFCFGATER--CGNSNLARHEFVDMEGWLLSSDRQVVPHEHPEIELLHAD 1682  
  
QY 165 -----DGCNNSWIPRGNYIESNRDDCTVSLIYAVHLK-KSGYVFFFEYQYVDNNIFFEFF 218  
Db 1683 LRSVADTFSELYWQAPPSYLGDRVSSYGGTLYHELHSETQRGDIFIPYESRPDVV----- 1737  
  
QY 219 IQNDQCQ-----EMDTTDKWKVL-TDNGEWGSHSVMLKSGTNILYWR 260  
Db 1738 LQGNQMSIAFLAYPPPGQVHRGQLQLVEGNFRHLETHNPVSRREELMMVLAGLEQLQIR 1797  
  
QY 261 TTGILMGSKAVKPVLYKNITIEGVAYT-----SECFPCPKPGTFSN 300  
Db 1798 A----LFSQTSSSVSLRRVVLVAVASEAGRGPPASNVELCMCPANVRGDSQCQECAPGYRD 1853  
  
QY 301 KPGSF--NCQVCPRNTYSEK---GAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTP 355  
Db 1854 TKGLFLGRCVPCQCHGHSRCLPLPGSGIGVCQHNTEGDQCERC--RPGFVSSDSPSNPASP 1911  
  
QY 356 CD-----EEGKTQIMYKWIEPKIKICREDLTDAIRLPPSGEKKDCPPC 396  
Db 1912 CVSCPCPLAVPSNNFADGCVLNRGTQC-----LCR-----PGYAGASCERC 1953  
  
QY 397 NPGFYNNG---SSSCHPC-----PGTFFSDG---TKECRPC-----PAGTEPALGFE 437  
Db 1954 APGFFGNPLVLGSSCQPCDCSGNGDPNMFSDCDPLTGACRGCLRHTTGPHCERCAPGF- 2012



QY 438 YKWN-VLPGN-MKTSCFNVGNSKCDGMNGWEVAGDHIOGAGGSDNDYLILNLHIPGFK 495  
Db 2013 --YGNALLPGNCTRCDCSPCGTETCDPSG----- 2040  
QY 496 PPTSMTCATGSELGRITTFVETLCSADCVLYFMVDINRKSTNVVSWGGTKEKQAYTHII 555  
Db 2041 -----RCLCKAGVT-----GQRCDRCLEGYFGFEOCQ----- 2067  
QY 556 FKNATFTFWAFORTNQDNRREFINDVMVKIYSITATNAVDGVASSCRACALGSEQSGSS 615  
Db 2068 -----GCRPCACGPAAGKSE 2082  
QY 616 CVP-----CPPGHY--IEKETNQCK-----ECPP----- 637  
Db 2083 CHPQSGQCHCQPGTGTGQCLECAPGYWGLPEKGCRRQCQPRGHCDPHTGHCTCPPGLSGE 2142  
QY 638 --DTYLSIHQVYGKEACIPCGPGSKNNQ---DHSVVCYSDCFYHEKENQILHYDFSMLS 691  
Db 2143 RCDTCSQHQV-----PVPKPGGHHGHIHCEVCDHCV-----VLLLD--DLE 2181  
QY 692 SVGSL-----MNGPSTSKGTKYFHFENISLCGHEGK 723  
Db 2182 RAGALLPAIREQLQGINASSAANRLHRLNASIADLQSK 2220  
RESULT 6  
ZAN\_MOUSE STANDARD; PRT; 5376 AA.  
ID ZAN\_MOUSE  
AC O88799; O08647;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Zonadhesin precursor.  
GN ZAN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=98123114; PubMed=9452463;  
RA Gao Z., Garbers D.L.;  
RT "Species diversity in the structure of zonadhesin, a sperm-specific  
RT membrane protein containing multiple cell adhesion molecule-like  
RT domains.";  
RL J. Biol. Chem. 273:3415-3421(1998).  
RN [2]  
RP SEQUENCE OF 4864-5376 FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=97271566; PubMed=9126492;  
RA Gao Z., Harumi T., Garbers D.L.;  
RT "Chromosome localization of the mouse zonadhesin gene and the human  
RT zonadhesin gene (ZAN).";  
RL Genomics 41:119-122(1997).  
CC -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA  
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR  
CC SIGNALING.  
CC -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE  
CC APICAL REGION OF THE SPERM HEAD.  
CC -!- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.  
CC -!- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATES SPERM ADHESION TO THE  
CC ZONA PELLUCIDA.  
CC -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,  
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF  
CC SPERMATOZOEA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.  
CC -!- DOMAIN: THE VWFD DOMAIN 2 MAY MEDIATE COVALENT  
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).  
CC -!- SIMILARITY: CONTAINS 3 MAM DOMAINS.  
CC -!- SIMILARITY: CONTAINS 25 VWFD DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
CC EMBL; U97068; AAC26680.1; -;  
CC EMBL; U83190; AAC53125.1; -;  
CC MGI; 106656; Zan.  
CC InterPro; IPR000561; EGF-like.  
CC InterPro; IPR003645; FOLN.  
CC InterPro; IPR000998; MAM\_domain.  
CC InterPro; IPR002919; TIL\_Cysrich.  
CC InterPro; IPR003328; TILA\_Cysrich.  
CC InterPro; IPR001007; VWF\_C.  
CC InterPro; IPR001846; VWF\_D.  
CC Pfam; PF00094; vwd; 4.  
CC Pfam; PF00629; MAM; 3.  
CC Pfam; PF01826; TIL; 25.  
CC Pfam; PF02345; TILA; 25.  
CC SMART; SM00181; EGF; 2.  
CC SMART; SM00274; FOLN; 11.  
CC SMART; SM00137; MAM; 2.  
CC SMART; SM00214; WVC; 17.  
CC SMART; SM00216; VWD; 4.  
CC PROSITE; PS00022; EGF\_1; 1.  
CC PROSITE; PS01186; EGF\_2; 18.  
CC PROSITE; PS00740; MAM\_1; FALSE\_NEG.  
CC PROSITE; PS00060; MAM\_2; 3.  
KW Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;  
KW Repeat.  
FT SIGNAL 1 17 POTENTIAL.  
FT CHAIN 18 5376 ZONADHESIN.  
FT DOMAIN 18 5310 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 5311 5337 POTENTIAL.  
FT DOMAIN 5338 5376 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 45 210 MAM 1.  
FT DOMAIN 215 374 MAM 2.  
FT DOMAIN 377 542 MAM 3.  
FT DOMAIN 547 1170 80 X HEPTAPEPTIDE REPEATS (APPROXIMATE)  
FT DOMAIN 1171 1280 (MUCIN-LIKE DOMAIN).  
FT DOMAIN 1281 1669 VWFD 1 (PARTIAL).  
FT DOMAIN 1670 2056 VWFD 2.  
FT DOMAIN 2057 2459 VWFD 3.  
FT DOMAIN 2460 2579 VWFD 4.  
FT DOMAIN 2580 2699 VWFD 5 (PARTIAL).  
FT DOMAIN 2700 2819 VWFD 6 (PARTIAL).  
FT DOMAIN 2820 2939 VWFD 7 (PARTIAL).  
FT DOMAIN 2940 3059 VWFD 8 (PARTIAL).  
FT DOMAIN 3060 3179 VWFD 9 (PARTIAL).  
FT DOMAIN 3180 3299 VWFD 10 (PARTIAL).  
FT DOMAIN 3300 3416 VWFD 11 (PARTIAL).  
FT DOMAIN 3417 3536 VWFD 12 (PARTIAL).  
FT DOMAIN 3537 3656 VWFD 13 (PARTIAL).  
FT DOMAIN 3657 3776 VWFD 14 (PARTIAL).  
FT DOMAIN 3777 3892 VWFD 15 (PARTIAL).  
FT DOMAIN 3893 4928 VWFD 16 (PARTIAL).  
FT DOMAIN 4029 4148 VWFD 17 (PARTIAL).  
FT DOMAIN 4149 4263 VWFD 18 (PARTIAL).  
FT DOMAIN 4264 4283 VWFD 19 (PARTIAL).  
FT DOMAIN 4384 4503 VWFD 20 (PARTIAL).  
FT DOMAIN 4504 4623 VWFD 21 (PARTIAL).  
FT DOMAIN 4624 4743 VWFD 22 (PARTIAL).  
FT DOMAIN 4744 4863 VWFD 23 (PARTIAL).  
FT DOMAIN 4864 5261 VWFD 24 (PARTIAL).  
FT DOMAIN 5259 5295 VWFD 25.  
FT DISULFID 5263 EGF-LIKE.  
FT DISULFID 5268 BY SIMILARITY.  
FT DISULFID 5285 BY SIMILARITY.  
FT CARBOHYD 339 BY SIMILARITY.  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
339



FT	CARBOHYD	499	499	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1216	1216	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1239	1239	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1314	1314	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1814	1814	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1908	1908	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1933	1933	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2028	2028	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2111	2111	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2142	2142	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2332	2332	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2533	2533	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2575	2575	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2692	2692	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2812	2812	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3052	3052	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3065	3065	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3144	3144	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3172	3172	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3288	3288	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3292	3292	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3782	3782	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	4005	4005	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	4136	4136	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	4243	4243	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	4254	4254	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	4335	4335	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	4376	4376	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	4586	4586	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	5136	5136	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	5252	5252	N-LINKED	(GLCNAC. . .)	(POTENTIAL).
SQ	SEQUENCE	5376 AA; 579908 MW; 0E44DB77DF2A2620 CRC64;				
Query Match 3.6%; Score 194; DB 1; Length 5376;						
Best Local Similarity 20.5%; Pred. No. 1.7e-05;						
Matches 144; Conservative 63; Mismatches 204; Indels 290; Gaps 38;						
Qy	75	SSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNOVCSCKCGEGTYSLGSGI	134			
Db	3256	SRGCTQSTCPAGAIHCR-----NFKCPSTGYCKNGDNGSSNCTEITLQCPTNS	3304			
Qy	135	KFDEWDELPAFGFSNIATFMDTVVGPDSRDPDGCNNSSWIPRGNVIESNRDDCTVSLIYAV	194			
Db	3305	QFT--DCLP-----SCVPSCSNRCEVTSVPSPSSCREGC-----	3336			
Qy	195	HLKKSQGVFFEXQYVDNNIFFEFTQDQCQEMDTTDDKVKVLTNDGEGWSHVMKLSGT	254			
Db	3337	-LCNHGVFSE-----DKCVPRTQCGCKDARGAIIPAG-	3368			
Qy	255	NILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECFPCPKPGTFSNKPGSFNCQ--VCPR	312			
Db	3369	-----KTWTSKGCTQSCACV-----EGNIQCQNFCQCPP	3396			
Qy	313	NTYSEKAKECIRCKDDSQFSGSECTE---RPPCTTKDYFQIHTPC-----DEEG	360			
Db	3397	ETY-----CKDNSE--GSSFTKITLQCPAHTQ-----YTSCLPCLPSCLDPEG	3439			
Qy	361	KTQIMYKWIETPKI---CRE-----DLTDAT-RLPPSGEKKDCPP	395			
Db	3440	---LCKDISPKVPSTCKEGVCQSGYVLNSDKCVLRAECDCKDAQGALIPAGKTTWTS	3495			
Qy	396	CNPG-FYNGSSSCH--PCPPGTF---SDGTKECR---PCPAGTEPALGFYKWNVL	444			
Db	3496	CTQSCACMGGAVQCQSSQCPGPTCKDNEDGNSNCAKITLQCPAHS-----LFTNCL	3547			
Qy	445	PGNMKTSFNVGNSKCDGMNGEVAGDHIQSGAGGSDNDYLILNLHIPGPKPPTSMTGAT	504			
Db	3548	PPCL-PSCLDP-----DGL-----CKGASPKVPSTCKE	3574			
Qy	505	GSELGRITVFETLCSADCVLYFMVDINRKSTN--VVESWGGFKQKAYTHIIFKNATFT	562			
Db	3575	G-----CICQSGYVL-----SNNKCLLRNRCCGKD--AHGALIPEDK---	3609			

QY	563	FTWAFQRTNQODNRRFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPG	622			
Db	3610	-TWVSR-----GCTQSC-VCTGSGSIQCLSS--QCPPG	3637			
QY	623	HYI---EKETNOCKECPDPTYLSIHQVYGKEACIPCGPGSKNNQDH-----SVCYS	670			
Db	3638	AYCKDNEDGSSNCARIPPQCPANSHYT---DCFPCCPPSCSDPEGHCEASGPRVLSTCRE	3694			
QY	671	DCF----FYHEKENQILHYDFSLSVSGSLM-NGPSFTSKG	706			
Db	3695	GCLCNPGFVLRDKCVPRECGCKDAQGALIPSGKTTWTS	3735			
RESULT 7						
TS11_GIALA						
ID	TS11_GIALA	STANDARD;	PRT;	667	AA.	
AC	Q03185;					
DT	01-JUN-1994 (Rel. 29, Created)					
DT	01-JUN-1994 (Rel. 29, Last sequence update)					
DT	15-JUL-1998 (Rel. 36, Last annotation update)					
DE	Major surface trophozoite antigen 11 precursor.					
GN	TS11.					
OS	Giardia lamblia (Giardia intestinalis).					
OC	Eukaryota; Diplomonadida; Hexamitidae; Giardiae; Giardia.					
OX	NCBI_TaxID=5741;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=Isolate AD-1;					
RX	MEDLINE=93241215; PubMed=8479449;					
RA	Ey P.L., Khanna K., Manning P.A., Mayrhofer G.;					
RT	"A gene encoding a 69-kilodalton major surface protein of Giardia intestinalis trophozoites.";					
RL	Mol. Biochem. Parasitol. 58:247-258(1993).					
CC	-!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE PLASMA MEMBRANE.					
CC	-!- DOMAIN: CONTAINS 27 REPEATS OF A CXXC MOTIF.					
CC	-!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).					
CC	EMBL; M95814; AAA02687.1; --					
DR	PIR; A48579; A48579.					
DR	HSSP; P02468; 1KLO.					
DR	InterPro; IPR000561; EGF-like.					
DR	InterPro; IPR002174; Furin-like.					
DR	InterPro; IPR005127; Giardia_VSP.					
DR	Pfam; PF03302; VSP; 1.					
DR	SMART; SM00181; EGF; 3.					
DR	SMART; SM00261; FU; 5.					
KW	Antigen; Repeat; Transmembrane; Signal.					
FT	SIGNAL	1	17			
FT	CHAIN	18	667			
FT	DOMAIN	18	633			
FT	TRANSMEM	634	662			
FT	DOMAIN	663	667			
FT	CARBOHYD	591	591			
FT	CARBOHYD	630	630			
SQ	SEQUENCE	667	AA; 68475	MW; 1DD9572703232B8D	CRC64;	
Query Match 3.4%; Score 182; DB 1; Length 667;						
Best Local Similarity 19.9%; Pred. No. 8.8e-06;						
Matches 146; Conservative 58; Mismatches 236; Indels 292; Gaps 40;						
QY	103	CTFSCASGEYLEMKNOVCSCGEG-TYSLGSGIKFDEWDELPAFGFSNIATFMDTVVGP	SD	161		
Db	141	CSAPKAPGEDNTPKAATCTKCAAGFLHTPSEGLSSCE-ETCPEGY-----FGHTATAESK	194			



```
FT DISULFID 447 456 BY SIMILARITY. (POTENTIAL).
FT DISULFID 459 469 BY SIMILARITY. (POTENTIAL).
FT DISULFID 472 484 BY SIMILARITY. (POTENTIAL).
FT DISULFID 474 491 BY SIMILARITY. (POTENTIAL).
FT DISULFID 493 502 BY SIMILARITY. (POTENTIAL).
FT DISULFID 505 516 BY SIMILARITY. (POTENTIAL).
FT DISULFID 519 531 BY SIMILARITY. (POTENTIAL).
FT DISULFID 521 538 BY SIMILARITY. (POTENTIAL).
FT DISULFID 540 549 BY SIMILARITY. (POTENTIAL).
FT DISULFID 552 561 BY SIMILARITY. (POTENTIAL).
FT DISULFID 564 576 BY SIMILARITY. (POTENTIAL).
FT DISULFID 566 583 BY SIMILARITY. (POTENTIAL).
FT DISULFID 585 594 BY SIMILARITY. (POTENTIAL).
FT DISULFID 597 607 BY SIMILARITY. (POTENTIAL).
FT DISULFID 610 622 BY SIMILARITY. (POTENTIAL).
FT DISULFID 612 629 BY SIMILARITY. (POTENTIAL).
FT DISULFID 631 640 BY SIMILARITY. (POTENTIAL).
FT DISULFID 643 653 BY SIMILARITY. (POTENTIAL).
FT DISULFID 656 668 BY SIMILARITY. (POTENTIAL).
FT DISULFID 658 674 BY SIMILARITY. (POTENTIAL).
FT DISULFID 676 685 BY SIMILARITY. (POTENTIAL).
FT DISULFID 688 698 BY SIMILARITY. (POTENTIAL).
FT DISULFID 701 715 BY SIMILARITY. (POTENTIAL).
FT DISULFID 703 724 BY SIMILARITY. (POTENTIAL).
FT DISULFID 726 735 BY SIMILARITY. (POTENTIAL).
FT DISULFID 738 753 BY SIMILARITY. (POTENTIAL).
FT DISULFID 756 770 BY SIMILARITY. (POTENTIAL).
FT DISULFID 758 777 BY SIMILARITY. (POTENTIAL).
FT DISULFID 779 788 BY SIMILARITY. (POTENTIAL).
FT DISULFID 791 806 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1415 1427 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1417 1434 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1436 1445 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1448 1458 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1461 1469 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1463 1476 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1478 1487 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1490 1503 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1506 1520 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1529 1538 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1541 1551 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1554 1566 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1556 1573 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1575 1584 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1587 1602 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1830 1839 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1832 1846 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1849 1858 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1861 1877 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1880 1894 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1882 1905 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1907 1916 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1919 1934 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1937 1951 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1939 1958 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1961 1970 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1973 1987 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1990 2000 BY SIMILARITY. (POTENTIAL).
FT DISULFID 1992 2007 BY SIMILARITY. (POTENTIAL).
FT DISULFID 2009 2018 BY SIMILARITY. (POTENTIAL).
FT DISULFID 2021 2034 BY SIMILARITY. (POTENTIAL).
FT DISULFID 2037 2048 BY SIMILARITY. (POTENTIAL).
FT DISULFID 2039 2055 BY SIMILARITY. (POTENTIAL).
FT DISULFID 2057 2066 BY SIMILARITY. (POTENTIAL).
FT DISULFID 2069 2081 BY SIMILARITY. (POTENTIAL).
FT DISULFID 2084 2096 BY SIMILARITY. (POTENTIAL).
FT DISULFID 2086 2103 BY SIMILARITY. (POTENTIAL).
FT DISULFID 2105 2114 BY SIMILARITY. (POTENTIAL).
FT DISULFID 2117 2129 BY SIMILARITY. (POTENTIAL).
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 511 511 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 634 634 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 761 761 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1014 1014 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1341 1341 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1705 1705 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1756 1756 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1868 1868 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1944 1944 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1986 1986 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2002 2002 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2159 2159 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2207 2207 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2231 2231 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2235 2235 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2401 2401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2421 2421 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2487 2487 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2821 2821 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3087 3087 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3242 3242 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3541 3541 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3672 AA; 404223 MW; 28E262DB5FF14BFA CRC64;

Query Match 3.3%; Score 178.5; DB 1; Length 3672;
Best Local Similarity 22.6%; Pred. No. 0.00014;
Matches 116; Conservative 54; Mismatches 203; Indels 141; Gaps 30;

QY 279 ITIEGVAYTSECFPCPKP--GTFESNKPQSF--NCQVCPRNTYSEKGAKECIRCDDSQFSG 334
: : || || || || : || || : | : | : || : | :
Db 1413 LSCDCVAQGSSEFQCEQYGGQCKCKPGVIGRRERCAPGYN---FPCEIKC----QCNA 1465

QY 335 SSECTERP-PCTTKDYFQIHTPCDECKTQIMYKWIEPKI-----CREDLTDAIRL--- 384
: || | : : | || : | : | : | : | : | : |
Db 1466 GQCDERTGQCFCPPHVEGQT-CDRCVSNAGY---DPLIGCQKCGCHPGSGEGNLVCD 1521

QY 385 PPSGE-----KKDCPPCPNPGFYNNGSSSCHPCP---PGT---FSDGTK----- 421
| || : : | || | : | : | : | : | : |
Db 1522 PESGQCLRESMGGRCQDRCLAGFY--GFPHCYGCSNRRAGTTEICDATNAQCKCKENV 1579

QY 422 ---ECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAG 478
| | || | : | : | : | : | : | : | : | : |
Db 1580 YGGRCEACKAGT-----FDLSAENPL-GCVNCFCFGVTDSCRSSMYVPVTIMSDMSSFLT 1633

QY 479 GSDN-----DYLILNLHIPG-----FKPPTSMTGATGSELGRITVFETL-----C 519
|| | : | : | : | : | : | : | : | : |
Db 1634 TDDNGMVDNKDDTVIYTSEETSPNSVYFNVPIEKKDYTTSYGLKLTFKLSTVPRGGRKSM 1693

QY 520 SADCVLYFMVDINRKSTNV-VESWGG---TKEKQAYTHIIFKNATFTFTWAFQRTNOQGD 575
: || : : | : | : | : | : | : | : | : | :
Db 1694 NAD-----ADVRLTGANMTIEYWASEQPTNPEEQFT-VKCKLVPENFLTAEGKTVTREE 1746

QY 576 NRRFINDMVKI-----YSITATNAVDBGV--ASSCRACALGSEQSGS 614
: : : | : | : | : | : | : | : | : | : |
Db 1747 LMKVLHSLQNITLKASYFDHPKTSLYEFGLEISEPNGVDSVIKASSVEQCQC PAPYTG 1806

QY 615 SCVPCPPGHYIEKETNQCKECPDPTVLSIHQVYKGE---ACIPC---GPGSKNNQDHSVC 668
|| | : : | : | : | : | : | : | : | : | :
Db 1807 SCQLCASGY-----HRVQSGSFLGACVPCPCNGHSATCDPDTGIC 1846

QY 669 YSDCFFYHEKENQILHYDFSNLSSVSGSLMNGPSF 702
: || | : | : | : | : | : | : | : | :
Db 1847 -TDC----EHTNGDHCFECNEGHYGNATNGSPY 1875

RESULT 9
LMGL_MOUSE
ID LMGL_MOUSE STANDARD; PRT; 1607 AA.
AC P02468;
DT 21-JUL-1986 (Rel. 01, Created)
```



DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Laminin gamma-1 chain precursor (Laminin B2 chain).  
GN LAMC1 OR LAMC-1 OR LAMB-2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88059118; PubMed=3680290;  
RA Sasaki M., Yamada Y.;  
RT "The laminin B2 chain has a multidomain structure homologous to the  
RT B1 chain."  
RL J. Biol. Chem. 262:17111-17117(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89000737; PubMed=3167041;  
RA Durkin M.E., Bartos B.B., Liu S.-H., Phillips S.L., Chung A.E.;  
RT "Primary structure of the mouse laminin B2 chain and comparison with  
RT laminin B1."  
RL Biochemistry 27:5198-5204(1988).  
RN [3]  
RP SEQUENCE OF 1-239 FROM N.A.  
RX MEDLINE=88228071; PubMed=2836421;  
RA Ogawa K., Burbelo P.D., Sasaki M., Yamada Y.;  
RT "The laminin B2 chain promoter contains unique repeat sequences and  
RT is active in transient transfection."  
RL J. Biol. Chem. 263:8384-8389(1988).  
RN [4]  
RP SEQUENCE OF 1391-1607 FROM N.A.  
RX MEDLINE=85051302; PubMed=6209134;  
RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;  
RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of  
RT coiled-coil alpha-helix."  
RL EMBO J. 3:2355-2362(1984).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 771-932.  
RX MEDLINE=96196434; PubMed=8648630;  
RA Stetefeld J., Mayer U., Timpl R., Huber R.;  
RT "Crystal structure of three consecutive laminin-type epidermal growth  
RT factor-like (LE) modules of laminin gammal chain harboring the  
RT nidogen binding site."  
RL J. Mol. Biol. 257:644-657(1996).  
RN [6]  
RP STRUCTURE BY NMR OF 824-881.  
RX MEDLINE=96196435; PubMed=8648631;  
RA Baumgartner R., Czisch M., Mayer U., Poeschl E., Huber R.,  
RA Timpl R., Holak T.A.;  
RT "Structure of the nidogen binding LE module of the laminin gammal  
RT chain in solution."  
RL J. Mol. Biol. 257:658-668(1996).  
CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin  
CC is thought to mediate the attachment, migration, and organization  
CC of cells into tissues during embryonic development by interacting  
CC with other extracellular matrix components.  
CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three  
CC different polypeptide chains (alpha, beta, gamma), which are bound  
CC to each other by disulfide bonds into a cross-shaped molecule  
CC comprising one long and three short arms with globules at each  
CC end.  
CC THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ),  
CC LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),  
CC LAMININ-6 (K-LAMININ) AND LAMININ-7 (KS-LAMININ).  
CC -!- SUBCELLULAR LOCATION: Extracellular.  
CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR  
CC COMPONENT).  
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT  
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.  
CC -!- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.  
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
CC -!- SIMILARITY: CONTAINS 11 LAMININ EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X05211; CAA28838.1; -;  
DR EMBL; J03484; AAA39405.1; -;  
DR EMBL; J02930; AAA39408.1; -;  
DR EMBL; J03749; AAA39409.1; -;  
DR PIR; A28469; MMSB2.  
DR PDB; 1KLO; 20-AUG-97.  
DR PDB; 1TLE; 12-FEB-97.  
DR MGD; MGI:99914; Lamc1.  
DR InterPro; IPR004089; Chmtaxis\_transd.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001886; LamNT.  
DR InterPro; IPR000034; Laminin\_B.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR Pfam; PF00052; laminin\_B; 1.  
DR Pfam; PF00053; laminin\_EGF; 10.  
DR Pfam; PF00055; laminin\_Nterm; 1.  
DR PRINTS; PR00011; EGFLAMININ.  
DR ProDom; PD002082; LamNT; 1.  
DR ProDom; PD003031; Laminin\_B; 1.  
DR SMART; SM00180; EGF\_Lam; 9.  
DR SMART; SM00001; EGF\_Like; 1.  
DR SMART; SM00281; LamB; 1.  
DR SMART; SM00136; LamNT; 1.  
DR PROSITE; PS00022; EGF\_1; 8.  
DR PROSITE; PS01186; EGF\_2; 2.  
DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 10.  
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.  
FT SIGNAL 1 33  
FT CHAIN 34 1607 LAMININ GAMMA-1 CHAIN.  
FT DOMAIN 34 283 LAMININ N-TERMINAL (DOMAIN VI).  
FT DOMAIN 284 339 LAMININ EGF-LIKE 1.  
FT DOMAIN 340 395 LAMININ EGF-LIKE 2.  
FT DOMAIN 396 442 LAMININ EGF-LIKE 3.  
FT DOMAIN 443 492 LAMININ EGF-LIKE 4.  
FT DOMAIN 493 502 LAMININ EGF-LIKE 5 (N-TERMINAL).  
FT DOMAIN 503 687 LAMININ DOMAIN IV.  
FT DOMAIN 688 721 LAMININ EGF-LIKE 5 (C-TERMINAL).  
FT DOMAIN 722 770 LAMININ EGF-LIKE 6.  
FT DOMAIN 771 825 LAMININ EGF-LIKE 7.  
FT DOMAIN 826 881 LAMININ EGF-LIKE 8 (NIDOGEN-BINDING).  
FT DOMAIN 882 932 LAMININ EGF-LIKE 9.  
FT DOMAIN 933 980 LAMININ EGF-LIKE 10.  
FT DOMAIN 981 1028 LAMININ EGF-LIKE 11.  
FT DOMAIN 1029 1607 DOMAIN II AND I.  
FT DOMAIN 1034 1594 COILED COIL (POTENTIAL).  
FT DISULFID 340 349 BY SIMILARITY.  
FT DISULFID 342 365 BY SIMILARITY.  
FT DISULFID 368 377 BY SIMILARITY.  
FT DISULFID 380 393 BY SIMILARITY.  
FT DISULFID 396 408 BY SIMILARITY.  
FT DISULFID 398 414 BY SIMILARITY.  
FT DISULFID 416 425 BY SIMILARITY.  
FT DISULFID 428 440 BY SIMILARITY.  
FT DISULFID 443 454 BY SIMILARITY.  
FT DISULFID 445 461 BY SIMILARITY.  
FT DISULFID 463 472 BY SIMILARITY.  
FT DISULFID 475 490 BY SIMILARITY.  
FT DISULFID 722 731 BY SIMILARITY.  
FT DISULFID 724 738 BY SIMILARITY.  
FT DISULFID 740 749 BY SIMILARITY.  
FT DISULFID 752 768 BY SIMILARITY.  
FT DISULFID 771 779 BY SIMILARITY.  
FT DISULFID 773 790

FT	DISULFID	793	802		
FT	DISULFID	805	823		
FT	DISULFID	826	840		
FT	DISULFID	828	847		
FT	DISULFID	850	859		
FT	DISULFID	862	879		
FT	DISULFID	882	896		
FT	DISULFID	884	903		
FT	DISULFID	905	914		
FT	DISULFID	917	930		
FT	DISULFID	933	945		
FT	DISULFID	935	952		
FT	DISULFID	954	963		
FT	DISULFID	966	978		
FT	DISULFID	981	993		
FT	DISULFID	983	999		
FT	DISULFID	1001	1010		
FT	DISULFID	1013	1026		
FT	DISULFID	1029	1029		
FT	DISULFID	1032	1032		
FT	DISULFID	1598	1598		
FT	CARBOHYD	58	58		
FT	CARBOHYD	132	132		
FT	CARBOHYD	574	574		
FT	CARBOHYD	648	648		
FT	CARBOHYD	1020	1020		
FT	CARBOHYD	1105	1105		
FT	CARBOHYD	1159	1159		
FT	CARBOHYD	1173	1173		
FT	CARBOHYD	1203	1203		
FT	CARBOHYD	1221	1221		
FT	CARBOHYD	1239	1239		
FT	CARBOHYD	1378	1378		
FT	CARBOHYD	1393	1393		
FT	CARBOHYD	1437	1437		
FT	CONFLICT	216	216		
FT	CONFLICT	260	260		
FT	CONFLICT	337	337		
FT	CONFLICT	447	448		
FT	CONFLICT	544	544		
FT	CONFLICT	662	662		
FT	CONFLICT	886	886		
FT	CONFLICT	1158	1158		
FT	CONFLICT	1434	1434		
FT	CONFLICT	1475	1475		
FT	CONFLICT	1576	1576		
SQ	SEQUENCE	1607 AA; 177297 MW; 81B7B08E4869F242 CRC64;			
Query Match 3.3%; Score 177.5; DB 1; Length 1607;					
Best Local Similarity 19.3%; Pred. No. 5.8e-05;					
Matches 215; Conservative 115; Mismatches 437; Indels 347; Gaps					
QY	13	GWGRPAEAPR-RGRSPWPSPAWICWALAGCAAWAGDLPSSSSRPLPPCQEKDYHFEYT	71		
Db	3	GGGRAALALQPRGR-----LWPLLAVALAAGCVRAAMD-----	36		
QY	72	ECDSSGSRWRVAIP---NSAVD-----CSGLPDPVRGKECTFSCASGEYLEMKNQVC	120		
Db	37	ECADGGRPQRCMPEFVNAAFNVTVVATNTCGTPPE---EYCVQTGVTG-----VTKSC	87		
QY	121	SKCGEGTSLGSGIKF--DEWDELPAGFSNIATFMDTVVGPSD-----	161		
Db	88	HLCDAGQQHLQHGAFLTDYNNQADTTWWQSOTMLAGVQVPNSINLTLHLGKAFDITYVR	147		
QY	162	-----SRPDGC-----NNSSWIP-----RGNYESNR-----DD-----CT---	187		
Db	148	LKFHTSRPESFAIYKRTREDGPIWIPYQYVYSGSCENTYSKANRGFIRTTGGDEQQAALCTDEF	207		
QY	188	--VSLIYAVHL-----KKSGYVFFEYQYVDNNIFFEFIQNDQCQEMDTTTDKWVKLT	238		
Db	208	SDISPLTGGNVAFSTLEGRPSAYNF-----DN-----SPVLOEWVTATDIRVTLN	252		
QY	239	DNGEWG-----SHSVMLKSGTNIYWRRTTGILMG-----SKAVKPVLVKNI-TIEGV	284		

Db	253	RLNTFGDEVFNEPKVLKS-----YYAISDFAVGGRCKNGHASECVKNEFDKLMCNCKHN	308		
QY	285	AYTSECFCPCPKGTFPSNKP-----GSFNCQVCPRNTYSEKGAKECIR	325		
Db	309	TYGVDCEKCLP-FFNDRPWRRAATAESASESLPCDCNCRSQECYFDPPELYSTGHGHCNTN	367		
QY	326	CKDDSQFSGSSECTER-----PPCTKDYFQIHTPCDEEGKTQIMYKWIIEPKICR	375		
Db	368	CRDNTDGAKCERENFFRLGNTACSPCHSPVGSLSLTCDSYGRCS-----CK	417		
QY	376	EDLTDAIRLPPSGEKKDCPPCNPNGFYNNNGSSSCHPCPPGTFSDGTKECR-----PCPAG	429		
Db	418	PGV-----MGDK--CDRCQPGFHSLTEAGCRPC--SCDLRGSTDECNVETGRCVCKDN	466		
QY	430	TEPALGFEYK-----WWNVLPGNMK--TSCFNVGNSK--CDGMNGWEV-----AG	470		
Db	467	VE---GFNCERCKPGFFNLESSNPKGCTPCFCFGHSSVCTNAVGSVYDISSTFQIDEDG	523		
QY	471	DHIQSGAG-----GSDND--YLILNLHIPG-FKPPTSMGTATGSELGR-ITFVF----	515		
Db	524	WRVEQRDGEASELWSSDRQDIAVISDSYFPRYFIAPVKFLGNQVLSYQNLSFSFRVDR	583		
QY	516	-ETLCSADCVLVFMVDINRKSTNVVESWGTKEKQAYTHIIFKNATFTFTW-----AF	567		
Db	584	RDTRLSAEDLVLEGAGL-RVSVPLIAQGNYSYPSETTVKYIFRLHEATDYPWRPALSPFEF	642		
QY	568	QRTNOGQDNRRFINDMVKI-----YSITATNAVDBGVASSCRACALGSEQSGSSCVPCPPG	622		
Db	643	Q-----KLLNNLTSIKIRGTYSSERTAGYLDVV-----TLQSRPG---PGVPA	682		
QY	623	HYIEKETNCKECPDPTYLSIHQVYGKEACIPCGPGSKNNQDHSVYSDCCFFY-----	675		
Db	683	TWVESCT-----CPVG-----YGGQFCETCLPGYRRRETPSLGPIYSPCVLCTCNHSE	729		
QY	676	-----HEKENQILHYDFSNNLSSVGLMNGPSETSKGTYFHFNFNLSL	717		
Db	730	TCDPETGVCDCRDNTAGPHCEKCKSDGIYGDSTLGTSSDCQPCPCPGSSCAIVPKTKEV	789		
QY	718	CGH-----EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFCVQSTIIPSESKGFRAAL	772		
Db	790	CTHCPTGTAGKRCELCDDGYFGDPL-----GSNGPVLRCRQCNDNIDPN-AVGNCNRL	843		
QY	773	SSQSIILADTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSTATTSINGRSTA	832		
Db	844	TGECLKCIYNTAGFYCD-----RCKEGFGFNPLAPN-----ADCKKACACNPNYGTQ	891		
QY	833	VKMRNPNPKSGAGVISVPSKCPAGTCDCGTFFYFLWESAEACPLCTEHDHFHEIGAC--KR	890		
Db	892	QQSSCNPVGTQCQCLPHVSGRDCGTCD--PGYYNLQSGQGCERCDCCHALGSTNGQCDIRT	949		
QY	891	GFQETLYVWNEPKWCIKGISLPEKKLATCETVDF	924		
Db	950	GQCE-----CQPGIT--GOHCERCETNHF	971		

RESULT 10  
FBN2\_MOUSE  
ID FBN2\_MOUSE STANDARD; PRT; 2907 AA.  
AC Q61555; Q63957;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Fibrillin 2 precursor.  
GN FBN2 OR FBN-2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheraia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95263670; PubMed=7744963;  
RA Zhang H., Hu W., Ramirez F.;

RT "Developmental expression of fibrillin genes suggests heterogeneity  
RT of extracellular microfibrils.";  
RL J. Cell Biol. 129:1165-1176(1995).  
RN [2]  
RP SEQUENCE OF 210-317 FROM N.A.  
RX MEDLINE=94140368; PubMed=8307578;  
RA Li X., Pereira L., Zhang H., Sanguinetti C., Ramirez F., Bonadio J.,  
RA Francke U.;  
RT "Fibrillin genes map to regions of conserved mouse/human synteny on  
RT mouse chromosomes 2 and 18.";  
RL Genomics 18:667-672(1993).  
CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS  
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE  
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.  
CC -!- SIMILARITY: CONTAINS 47 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L39790; AAA74908.1; -;  
DR EMBL; S69359; AAC60685.1; -;  
DR HSSP; P35555; IEMN.  
DR MGD; MGI:95490; Fbn2.  
DR InterPro; IPR000152; Asx\_hydroxyl.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR001438; EGF\_II.  
DR InterPro; IPR002212; Fibril-assoc.  
DR Pfam; PF00008; EGF; 46.  
DR Pfam; PF00683; TB; 9.  
DR PRINTS; PR00010; EGFBL00D.  
DR SMART; SM00179; EGF\_CA; 43.  
DR SMART; SM00001; EGF\_like; 3.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 43.  
DR PROSITE; PS00022; EGF\_1; 2.  
DR PROSITE; PS01186; EGF\_2; 36.  
DR PROSITE; PS01187; EGF\_CA; 43.  
KW Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;  
KW Repeat; Signal; Multigene family.  
FT SIGNAL 1 28 POTENTIAL.  
FT CHAIN 29 2907 FIBRILLIN 2.  
FT DOMAIN 111 142 EGF-LIKE 1.  
FT DOMAIN 145 176 EGF-LIKE 2.  
FT DOMAIN 176 208 EGF-LIKE 3.  
FT DOMAIN 276 317 EGF-LIKE 4, CALCIUM-BINDING.  
FT DOMAIN 318 359 EGF-LIKE 5, CALCIUM-BINDING.  
FT REPEAT 360 426 TGFBP 1.  
FT DOMAIN 487 527 EGF-LIKE 6.  
FT DOMAIN 528 567 EGF-LIKE 7, CALCIUM-BINDING.  
FT DOMAIN 568 609 EGF-LIKE 8, CALCIUM-BINDING.  
FT DOMAIN 610 650 EGF-LIKE 9, CALCIUM-BINDING.  
FT DOMAIN 651 691 EGF-LIKE 10, CALCIUM-BINDING.  
FT REPEAT 692 760 TGFBP 2.  
FT DOMAIN 761 802 EGF-LIKE 11, CALCIUM-BINDING.  
FT DOMAIN 803 844 EGF-LIKE 12, CALCIUM-BINDING.  
FT DOMAIN 845 883 EGF-LIKE 13, CALCIUM-BINDING.  
FT DOMAIN 948 989 EGF-LIKE 14, CALCIUM-BINDING.  
FT REPEAT 990 1065 TGFBP 3.  
FT DOMAIN 1066 1107 EGF-LIKE 15, CALCIUM-BINDING.  
FT DOMAIN 1108 1150 EGF-LIKE 16, CALCIUM-BINDING.  
FT DOMAIN 1151 1192 EGF-LIKE 17, CALCIUM-BINDING.  
FT DOMAIN 1193 1234 EGF-LIKE 18, CALCIUM-BINDING.  
FT DOMAIN 1235 1275 EGF-LIKE 19, CALCIUM-BINDING.  
FT DOMAIN 1276 1317 EGF-LIKE 20, CALCIUM-BINDING.  
FT DOMAIN 1318 1359 EGF-LIKE 21, CALCIUM-BINDING.  
FT DOMAIN 1360 1400 EGF-LIKE 22, CALCIUM-BINDING.  
FT DOMAIN 1401 1441 EGF-LIKE 23, CALCIUM-BINDING.

FT DOMAIN 1442 1483 EGF-LIKE 24, CALCIUM-BINDING.  
FT DOMAIN 1484 1524 EGF-LIKE 25, CALCIUM-BINDING.  
FT DOMAIN 1525 1565 EGF-LIKE 26, CALCIUM-BINDING.  
FT REPEAT 1566 1642 TGFBP 4.  
FT DOMAIN 1643 1684 EGF-LIKE 27, CALCIUM-BINDING.  
FT DOMAIN 1685 1726 EGF-LIKE 28, CALCIUM-BINDING.  
FT REPEAT 1727 1800 TGFBP 5.  
FT DOMAIN 1801 1842 EGF-LIKE 29, CALCIUM-BINDING.  
FT DOMAIN 1843 1884 EGF-LIKE 30, CALCIUM-BINDING.  
FT DOMAIN 1885 1926 EGF-LIKE 31, CALCIUM-BINDING.  
FT DOMAIN 1927 1965 EGF-LIKE 32, CALCIUM-BINDING.  
FT DOMAIN 1966 2008 EGF-LIKE 33, CALCIUM-BINDING.  
FT DOMAIN 2009 2048 EGF-LIKE 34, CALCIUM-BINDING.  
FT DOMAIN 2049 2090 EGF-LIKE 35, CALCIUM-BINDING.  
FT REPEAT 2091 2163 TGFBP 6.  
FT DOMAIN 2164 2205 EGF-LIKE 36, CALCIUM-BINDING.  
FT DOMAIN 2206 2245 EGF-LIKE 37, CALCIUM-BINDING.  
FT DOMAIN 2246 2286 EGF-LIKE 38, CALCIUM-BINDING.  
FT DOMAIN 2287 2330 EGF-LIKE 39, CALCIUM-BINDING.  
FT DOMAIN 2331 2372 EGF-LIKE 40, CALCIUM-BINDING.  
FT REPEAT 2373 2441 TGFBP 7.  
FT DOMAIN 2442 2483 EGF-LIKE 41, CALCIUM-BINDING.  
FT DOMAIN 2484 2524 EGF-LIKE 42, CALCIUM-BINDING.  
FT DOMAIN 2525 2563 EGF-LIKE 43, CALCIUM-BINDING.  
FT DOMAIN 2564 2606 EGF-LIKE 44, CALCIUM-BINDING.  
FT DOMAIN 2607 2646 EGF-LIKE 45, CALCIUM-BINDING.  
FT DOMAIN 2647 2687 EGF-LIKE 46, CALCIUM-BINDING.  
FT DOMAIN 2688 2727 EGF-LIKE 47, CALCIUM-BINDING.  
FT DISULFID 115 124 BY SIMILARITY.  
FT DISULFID 119 130 BY SIMILARITY.  
FT DISULFID 132 141 BY SIMILARITY.  
FT DISULFID 149 159 BY SIMILARITY.  
FT DISULFID 153 164 BY SIMILARITY.  
FT DISULFID 166 175 BY SIMILARITY.  
FT DISULFID 180 190 BY SIMILARITY.  
FT DISULFID 184 196 BY SIMILARITY.  
FT DISULFID 198 207 BY SIMILARITY.  
FT DISULFID 280 292 BY SIMILARITY.  
FT DISULFID 287 301 BY SIMILARITY.  
FT DISULFID 303 316 BY SIMILARITY.  
FT DISULFID 322 334 BY SIMILARITY.  
FT DISULFID 329 343 BY SIMILARITY.  
FT DISULFID 345 358 BY SIMILARITY.  
FT DISULFID 491 503 BY SIMILARITY.  
FT DISULFID 498 512 BY SIMILARITY.  
FT DISULFID 514 526 BY SIMILARITY.  
FT DISULFID 532 542 BY SIMILARITY.  
FT DISULFID 537 551 BY SIMILARITY.  
FT DISULFID 553 566 BY SIMILARITY.  
FT DISULFID 572 584 BY SIMILARITY.  
FT DISULFID 579 593 BY SIMILARITY.  
FT DISULFID 595 608 BY SIMILARITY.  
FT DISULFID 614 625 BY SIMILARITY.  
FT DISULFID 620 634 BY SIMILARITY.  
FT DISULFID 636 649 BY SIMILARITY.  
FT DISULFID 655 666 BY SIMILARITY.  
FT DISULFID 661 675 BY SIMILARITY.  
FT DISULFID 677 690 BY SIMILARITY.  
FT DISULFID 765 777 BY SIMILARITY.  
FT DISULFID 772 786 BY SIMILARITY.  
FT DISULFID 788 801 BY SIMILARITY.  
FT DISULFID 807 819 BY SIMILARITY.  
FT DISULFID 814 828 BY SIMILARITY.  
FT DISULFID 830 843 BY SIMILARITY.  
FT DISULFID 849 859 BY SIMILARITY.  
FT DISULFID 854 868 BY SIMILARITY.  
FT DISULFID 870 883 BY SIMILARITY.  
FT DISULFID 952 964 BY SIMILARITY.  
FT DISULFID 959 973 BY SIMILARITY.  
FT DISULFID 975 988 BY SIMILARITY.  
FT DISULFID 1070 1082 BY SIMILARITY.  
FT DISULFID 1077 1091 BY SIMILARITY.  
FT DISULFID 1093 1106 BY SIMILARITY.



FT	DISULFID	1112	1124	BY SIMILARITY.
FT	DISULFID	1119	1133	BY SIMILARITY.
FT	DISULFID	1135	1149	BY SIMILARITY.
FT	DISULFID	1155	1167	BY SIMILARITY.
FT	DISULFID	1162	1176	BY SIMILARITY.
FT	DISULFID	1178	1191	BY SIMILARITY.
FT	DISULFID	1197	1209	BY SIMILARITY.
FT	DISULFID	1204	1218	BY SIMILARITY.
FT	DISULFID	1220	1233	BY SIMILARITY.
FT	DISULFID	1239	1250	BY SIMILARITY.
FT	DISULFID	1246	1259	BY SIMILARITY.
FT	DISULFID	1261	1274	BY SIMILARITY.
FT	DISULFID	1280	1292	BY SIMILARITY.
FT	DISULFID	1287	1301	BY SIMILARITY.
FT	DISULFID	1303	1316	BY SIMILARITY.
FT	DISULFID	1322	1334	BY SIMILARITY.
FT	DISULFID	1329	1343	BY SIMILARITY.
FT	DISULFID	1345	1358	BY SIMILARITY.
FT	DISULFID	1364	1377	BY SIMILARITY.
FT	DISULFID	1371	1386	BY SIMILARITY.
FT	DISULFID	1388	1399	BY SIMILARITY.
FT	DISULFID	1405	1418	BY SIMILARITY.
FT	DISULFID	1412	1427	BY SIMILARITY.
FT	DISULFID	1429	1440	BY SIMILARITY.
FT	DISULFID	1446	1458	BY SIMILARITY.
FT	DISULFID	1453	1467	BY SIMILARITY.
FT	DISULFID	1469	1482	BY SIMILARITY.
FT	DISULFID	1488	1499	BY SIMILARITY.
FT	DISULFID	1494	1508	BY SIMILARITY.
FT	DISULFID	1510	1523	BY SIMILARITY.
FT	DISULFID	1529	1540	BY SIMILARITY.
FT	DISULFID	1535	1549	BY SIMILARITY.
FT	DISULFID	1551	1564	BY SIMILARITY.
FT	DISULFID	1647	1659	BY SIMILARITY.
FT	DISULFID	1654	1668	BY SIMILARITY.
FT	DISULFID	1670	1683	BY SIMILARITY.
FT	DISULFID	1689	1701	BY SIMILARITY.
FT	DISULFID	1696	1710	BY SIMILARITY.
FT	DISULFID	1712	1725	BY SIMILARITY.
Query Match 3.3%; Score 177.5; DB 1; Length 2907;				
Best Local Similarity 20.1%; Pred. No. 0.00012;				
Matches 224; Conservative 101; Mismatches 375; Indels 417; Gaps 67;				
QY	41 GCQAAWAGDL-----PSSSRRLPPCQEKDYHFEYTECDSSGSRW-VAIPNSAVDC 91			
Db	17 GCVALWAQGTGQPPPPKTLWPQPPQVPAVAGSEGGFMGPEYRDEGAVRASVR 76			
QY	92 SGLPDPVRGKE-C-----TFSCASGEYLEMKNQ-----VC-SKCGEGTYS----- 129			
Db	77 RGQQEILRGPNVCGSRFHSYCCPGWKTLPGGNQCIVIPICRNSCGDGFCSRPNMCTCSSGQ 136			
QY	130 -----LGSIGKFEDEWDELPAFESNIATFMDTVVCPDSR-DCGNN 169			
Db	137 ISPTCGRKSIQCCSVRCMNGGTCAADHCQCQKGY-----IGTYCGQPVCEGCGQN 186			
QY	170 SSWIPRGNVIESNRDDCTVSLIYAVHLKKSQYVF----FEYQYVDNNIFFEFTQNDCCQ 225			
Db	187 G-----GRCIGPNRCACV-----YGFTGPQCCERDVRTGPCFTQ--VNNQMCQ 226			
QY	226 EMDT---TTDKWKVKLTDNGEWGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIE 282			
Db	227 GOLTGIVCTKTLCCATIGRAWGHPCEMCPA-----QPQPCRPGFIPNIRTG 272			
QY	283 GVAYTSEC--FP--CKPGTFSNKPQSFNCQVCPNRTYSEKGAKECIRCKDDSQFSGSSEC 338			
Db	273 ACQDVDECOAIPGLCQGGNCINTVGSFECR-CP-----AGHKQ-----SET 312			
QY	339 TERPPCTTKDYFQIHTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNP 398			
Db	313 TQK-----C-EDIDECSVIPGVCETGDC----- 334			
QY	399 GFYNN--GSSSCHPCPPG--TFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKT----- 451			

Db	335 ---SNTVGSYFC-LCPRGFVTSTGSR-CIDQAGTCFSLVNGRCAQELPGRMAKAQCC 389			
QY	452 -----CFNVGN--SKC-----DGMNGWEVAGDHIQ--SGAGGSDNDYLILN- 488			
Db	390 CEPGRCSIGTIPEACPVRGSEYRRRLCLDGLPMGGIPGSSVSRPGGTGSGNGYGGGT 449			
QY	489 --LHIP---GFKPPTSMGTATGSELGRI---TFVFET-----LC----- 519			
Db	450 GFLPIPGDNGFSPGVGGAGVGAGGQGPITILNQTIDICKHHANLCLNGRCIPTVSS 509			
QY	520 -----SADCVLYFMVDINRKSTNVVESWGGTKEQAY---THIFKNATFTF 563			
Db	510 YRRCNMGYKODANGDC-----IDVDECTSNPCSDGCVNTPGSYCKCHA----- 555			
QY	564 TWAFQRTNQG-----DNRREFINDMVKIYSI-----TATNAVD-GV 598			
Db	556 --GFQRTPTKQACIDIDEQIONGVLCNKGRCVNSDGSFQICNAGFELTDGKNVCVDHDE 613			
QY	599 ASSCRACALG---SEQSGSSCVPCPPGHYTEKTNQC---KEC-PPDTYLSIHQV--YGK 649			
Db	614 CTTTNMCLNGMCINEDGSFKCV-CKPGFILAPNGRYCTDVDECQTPGICMNGHCINNEGS 672			
QY	650 EACIPCGPGSKNNQDHSVY-----SDCFYHEKENQILHYDFSNLSSVGSMLMGPSTF-- 702			
Db	673 FRC-DCPPGLAVGVDGRVVCVDTHMRSTCYGEIKKGVCVRPFPGAVTKSECCCAN-PDYGF 730			
QY	703 -----TSKGTQYFH-----FFNISLCGHEGKKMAL---CTNNITDFTVKEIVAGSDD 746			
Db	731 GPCQPCPARNSAEFHGLCSSGIGITVDGRDINECALDPDICANGICE----- 778			
QY	747 YTNLVGAFVCQSTIIPSESKGFRAALSSQSILADTFIGVTVETTLKNINIKED----- 800			
Db	779 --NLRGTYRCNC-----NSGYEPDASGRNCIDIDECL---VNRLLCDNGLCRNTPGSYS 827			
QY	801 -----MFPVPTSQIPDVHFFYKSSATTSCINGRST----AVKMRCNPTK--SGAGVI 847			
Db	828 CTCPPGYVLPETETCEDV-----NECESNPCVNGACRNNLGSFHCESPGSKLSSTGLI 882			
QY	848 SVPSKCPAGTC-----DGCTFYF-----LWESAEACPLCTEHDFHEIEG 886			
Db	883 CIDSL--KGTCLWLNQDNRCENVNINGATLKSECCATLGGAWGS--PCERC-----ELDA 932			
QY	887 ACKRGFQETLYVWNEPKWCIKGISLPEKKLATCETVD 923			
Db	933 ACPRGFAR-----IKGV-----TCEDVN 950			
RESULT 11				
LMG3_MOUSE				
ID	LMG3_MOUSE	STANDARD;	PRT;	1581 AA.
AC	Q9R0B6; Q9WTW6;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Laminin gamma-3 chain precursor (Laminin 12 gamma 3).			
GN	LAMC3.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Albus A.M., Burgeson B., Champlaud M.-F., Koch M., Olson P.;			
.RT	"Mouse laminin 12 gamma 3 chain.";			
RL	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 1-1526 FROM N.A.			
RX	MEDLINE=99253969; Pubmed=10318827;			
RA	Iivanainen A., Morita T., Tryggvason K.;			
RT	"Molecular cloning and tissue-specific expression of a novel murine			
RT	laminin gamma3 chain.";			
RL	J. Biol. Chem. 274:14107-14111(1999).			



QY 615 SCVPCPPGHY---IEKETNOCKECP-----PDTYLSIHQVYGKEACIPCGPGSKN 661  
Db 744 SCERCMPGFYGNFASGRADDCQPCPCQSQACATIPES-----GDIVVCTHCPPGQRG 795  
QY 662 NQDHSVVCYSDCFFYHEKENQILHYDFSLSVSGSLMNGPSTSKGTYFHFNFNISLCGHE 721  
Db 796 RRCS-C-EDGFF-----GDPLGLSGAP-----Q 817  
QY 722 GKKMALCTNITDFTVKEIVAGSDDYT-----NLVGAFCVQSTIIPSESKGFRAALS 773  
Db 818 PCRRQCQSGNVD---LNAVGNCDPHSGHCLRCLYNTTGAH-CEH-----857  
QY 774 SQSIILADFFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSTATTSCINGRSTAV 833  
Db 858 -----CREGRFGSAVATR-----PV-----DKCAPCSC-DLRGSVS 887  
QY 834 KMRCNPTKSGAGVISVP-----SKCPACTCDGCTFYFLWESAECPLCTEHDFHEIEG 886  
Db 888 EKTCPNPTGQC--VCLPYVSGRDCSRCSPG-----FYDL-QSGRCQCKCHPLGSLN 938  
QY 887 AC 888  
Db 939 KC 940

RESULT 12  
FBN1\_MOUSE STANDARD; PRT; 2871 AA.  
AC Q61554; Q60826;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Fibrillin 1 precursor.  
GN FBN1 OR FBN-1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95130561; PubMed=7829516;  
RA Yin W., Germiller J., Sanguinetti C., Smiley E., Pangilinan T.,  
Pereira L., Ramirez F., Bonadio J.;  
RT "Primary structure and developmental expression of Fbn-1, the mouse  
fibrillin gene.";  
RL J. Biol. Chem. 270:1798-1806(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD-1; TISSUE=Kidney;  
RA Ota K., Kumar A., Wada J., Liu Z., Kanwar Y.S.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS  
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE  
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.  
CC -!- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER  
CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE  
CC MICROFIBRILS (BY SIMILARITY).  
CC -!- SIMILARITY: CONTAINS 47 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L29454; AAA56840.1; -.  
DR EMBL; U22493; AAA64217.1; -.  
DR HSSP; P35555; LAPJ.  
DR MGD; MGI:95489; Fbn1.  
InterPro; IPR000152; Asx\_hydroxyl.

DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR002212; Fibril-assoc.  
DR Pfam; PF00008; EGF; 46.  
DR Pfam; PF00683; TB; 9.  
DR SMART; SM00179; EGF\_CA; 42.  
DR SMART; SM00001; EGF\_Like; 4.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 43.  
DR PROSITE; PS00022; EGF\_1; 2.  
DR PROSITE; PS01186; EGF\_2; 38.  
DR PROSITE; PS01187; EGF\_CA; 43.  
KW Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;  
KW Repeat; Signal; Multigene family.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 2871 FIBRILLIN 1.  
FT DOMAIN 81 112 EGF-LIKE 1.  
FT DOMAIN 115 146 EGF-LIKE 2.  
FT DOMAIN 147 178 EGF-LIKE 3.  
FT DOMAIN 246 287 EGF-LIKE 4.  
FT DOMAIN 288 329 EGF-LIKE 5.  
FT DOMAIN 330 401 TGFBP 1.  
FT DOMAIN 402 446 PRO-RICH.  
FT DOMAIN 449 489 EGF-LIKE 6.  
FT DOMAIN 490 529 EGF-LIKE 7.  
FT DOMAIN 530 571 EGF-LIKE 8.  
FT DOMAIN 572 612 EGF-LIKE 9.  
FT DOMAIN 613 653 EGF-LIKE 10.  
FT DOMAIN 656 721 TGFBP 2.  
FT DOMAIN 723 764 EGF-LIKE 11.  
FT DOMAIN 765 806 EGF-LIKE 12.  
FT DOMAIN 807 846 EGF-LIKE 13.  
FT DOMAIN 910 951 EGF-LIKE 14.  
FT DOMAIN 952 1018 TGFBP 3.  
FT DOMAIN 1028 1069 EGF-LIKE 15.  
FT DOMAIN 1070 1112 EGF-LIKE 16.  
FT DOMAIN 1113 1154 EGF-LIKE 17.  
FT DOMAIN 1155 1196 EGF-LIKE 18.  
FT DOMAIN 1197 1237 EGF-LIKE 19.  
FT DOMAIN 1238 1279 EGF-LIKE 20.  
FT DOMAIN 1280 1321 EGF-LIKE 21.  
FT DOMAIN 1322 1362 EGF-LIKE 22.  
FT DOMAIN 1363 1403 EGF-LIKE 23.  
FT DOMAIN 1404 1445 EGF-LIKE 24.  
FT DOMAIN 1446 1486 EGF-LIKE 25.  
FT DOMAIN 1487 1527 EGF-LIKE 26.  
FT DOMAIN 1528 1599 TGFBP 4.  
FT DOMAIN 1606 1647 EGF-LIKE 27.  
FT DOMAIN 1648 1688 EGF-LIKE 28.  
FT DOMAIN 1689 1758 TGFBP 5.  
FT DOMAIN 1766 1807 EGF-LIKE 29.  
FT DOMAIN 1808 1848 EGF-LIKE 30.  
FT DOMAIN 1849 1890 EGF-LIKE 31.  
FT DOMAIN 1891 1929 EGF-LIKE 32.  
FT DOMAIN 1930 1972 EGF-LIKE 33.  
FT DOMAIN 1973 2012 EGF-LIKE 34.  
FT DOMAIN 2013 2054 EGF-LIKE 35.  
FT DOMAIN 2055 2121 TGFBP 6.  
FT DOMAIN 2127 2165 EGF-LIKE 36.  
FT DOMAIN 2166 2205 EGF-LIKE 37.  
FT DOMAIN 2206 2246 EGF-LIKE 38.  
FT DOMAIN 2247 2290 EGF-LIKE 39.  
FT DOMAIN 2291 2332 EGF-LIKE 40.  
FT DOMAIN 2333 2400 TGFBP 7.  
FT DOMAIN 2402 2443 EGF-LIKE 41.  
FT DOMAIN 2444 2484 EGF-LIKE 42.  
FT DOMAIN 2485 2523 EGF-LIKE 43.  
FT DOMAIN 2524 2566 EGF-LIKE 44.  
FT DOMAIN 2567 2606 EGF-LIKE 45.  
FT DOMAIN 2607 2647 EGF-LIKE 46.  
FT DOMAIN 2648 2687 EGF-LIKE 47.  
FT DISULFID 85 94 BY SIMILARITY.  
FT DISULFID 89 100 BY SIMILARITY.  
FT DISULFID 102 111 BY SIMILARITY.



FT	DISULFID	119	129	BY SIMILARITY
FT	DISULFID	123	134	BY SIMILARITY
FT	DISULFID	136	145	BY SIMILARITY
FT	DISULFID	150	160	BY SIMILARITY
FT	DISULFID	154	166	BY SIMILARITY
FT	DISULFID	168	177	BY SIMILARITY
FT	DISULFID	250	262	BY SIMILARITY
FT	DISULFID	257	271	BY SIMILARITY
FT	DISULFID	273	286	BY SIMILARITY
FT	DISULFID	292	304	BY SIMILARITY
FT	DISULFID	299	313	BY SIMILARITY
FT	DISULFID	315	328	BY SIMILARITY
FT	DISULFID	453	465	BY SIMILARITY
FT	DISULFID	460	474	BY SIMILARITY
FT	DISULFID	476	488	BY SIMILARITY
FT	DISULFID	494	504	BY SIMILARITY
FT	DISULFID	499	513	BY SIMILARITY
FT	DISULFID	515	528	BY SIMILARITY
FT	DISULFID	534	546	BY SIMILARITY
FT	DISULFID	541	555	BY SIMILARITY
FT	DISULFID	557	570	BY SIMILARITY
FT	DISULFID	576	587	BY SIMILARITY
FT	DISULFID	582	596	BY SIMILARITY
FT	DISULFID	598	611	BY SIMILARITY
FT	DISULFID	617	628	BY SIMILARITY
FT	DISULFID	623	637	BY SIMILARITY
FT	DISULFID	639	652	BY SIMILARITY
FT	DISULFID	727	739	BY SIMILARITY
FT	DISULFID	734	748	BY SIMILARITY
FT	DISULFID	750	763	BY SIMILARITY
FT	DISULFID	769	781	BY SIMILARITY
FT	DISULFID	776	790	BY SIMILARITY
FT	DISULFID	792	805	BY SIMILARITY
FT	DISULFID	811	821	BY SIMILARITY
FT	DISULFID	816	830	BY SIMILARITY
FT	DISULFID	832	845	BY SIMILARITY
FT	DISULFID	914	926	BY SIMILARITY
FT	DISULFID	921	935	BY SIMILARITY
FT	DISULFID	937	950	BY SIMILARITY
FT	DISULFID	1032	1044	BY SIMILARITY
FT	DISULFID	1039	1053	BY SIMILARITY
FT	DISULFID	1055	1068	BY SIMILARITY
FT	DISULFID	1074	1086	BY SIMILARITY
FT	DISULFID	1081	1095	BY SIMILARITY
FT	DISULFID	1097	1111	BY SIMILARITY
FT	DISULFID	1117	1129	BY SIMILARITY
FT	DISULFID	1124	1138	BY SIMILARITY
FT	DISULFID	1140	1153	BY SIMILARITY
FT	DISULFID	1159	1171	BY SIMILARITY
FT	DISULFID	1166	1180	BY SIMILARITY
FT	DISULFID	1182	1195	BY SIMILARITY
FT	DISULFID	1201	1212	BY SIMILARITY
FT	DISULFID	1208	1221	BY SIMILARITY
FT	DISULFID	1223	1236	BY SIMILARITY
FT	DISULFID	1242	1254	BY SIMILARITY
FT	DISULFID	1249	1263	BY SIMILARITY
FT	DISULFID	1265	1278	BY SIMILARITY
FT	DISULFID	1284	1296	BY SIMILARITY
FT	DISULFID	1291	1305	BY SIMILARITY
FT	DISULFID	1307	1320	BY SIMILARITY
FT	DISULFID	1326	1339	BY SIMILARITY
FT	DISULFID	1333	1348	BY SIMILARITY
FT	DISULFID	1350	1361	BY SIMILARITY
FT	DISULFID	1367	1380	BY SIMILARITY
FT	DISULFID	1374	1389	BY SIMILARITY
FT	DISULFID	1391	1402	BY SIMILARITY
FT	DISULFID	1408	1420	BY SIMILARITY
FT	DISULFID	1415	1429	BY SIMILARITY
FT	DISULFID	1431	1444	BY SIMILARITY
FT	DISULFID	1450	1461	BY SIMILARITY
FT	DISULFID	1456	1470	BY SIMILARITY
FT	DISULFID	1472	1485	BY SIMILARITY
FT	DISULFID	1491	1502	BY SIMILARITY

FT	DISULFID	1497	1511	BY SIMILARITY.		
FT	DISULFID	1513	1526	BY SIMILARITY.		
FT	DISULFID	1610	1622	BY SIMILARITY.		
FT	DISULFID	1617	1631	BY SIMILARITY.		
FT	DISULFID	1633	1646	BY SIMILARITY.		
FT	DISULFID	1652	1663	BY SIMILARITY.		
FT	DISULFID	1658	1672	BY SIMILARITY.		
FT	DISULFID	1674	1687	BY SIMILARITY.		
Query Match 3.2%; Score 173.5; DB 1; Length 2871;						
Best Local Similarity 19.4%; Pred. No. 0.00024;						
Matches 216; Conservative 111; Mismatches 342; Indels 443; Gaps						
QY	35	CCWALAGCQA	AW-----AGDL	PSSSRPLPPCQEKDYHFEYTECDSSGSRWRVA-IPNSAV 89		
DB	980	CCCSVG--	AAWGTECEECPL	NSR-----EYEELCPRPGFATKDTNGKP 1024		
QY	90	-----DCSGLPDPVR	GKEC-----TFSCA--	SGPYLEMKNQVCSK-----CGEG 126		
DB	1025	FFKDINECKMIP	SLCTHCKRCRNTIGSFKRCDCSGFALDSEERNCTDIDECRISPLDCGRG 1084			
QY	127	TYSLGSG--	IKFDEWDELPA	GFNSIATEMDTVVGP	SD---SRPDGCNNSSW-----IPR 175	
DB	1085	QCVNTPGDFECK	CDEGYE--SGEMMKNCMDIDECQDP	LLCRGGGICHNTGTYRCECPP 1142		
QY	176	GNYIESN-----	RDDCTVSLIYAVHLK	KSGYVFEFYQYVDNNIFE-----FFIQNDQC 224		
DB	1143	GHOLSPNISACID	INECELSANLCPH--GRCVN	LIGKYQCACNPGYHPHPTHDR	LFRCVDIDEC 1201	
QY	225	QEMDTTTDKWV	KLTDNGEWGSHSVMI	KSCTNILYWRTTGILMGSKAVKPV	LKNTIEGV 284	
DB	1202	SIMNGGCET	FTCTNSD---GSY-----	-----	1219	
QY	285	AYTSECFPCPK	PGTFSNKPGSFNCVQ	CPRNTYSEKGAKECIRCKDD	SQFSGSSECTERP-- 342	
DB	1220	---EC-SCQPG-	FALMPDQ	RS-----TDIDQCEDNPNICDGGQCTNIPGE 1260		
QY	343	-PCTTKDYFQI	HTPCDEEGKTQIMYKW--	IEPKICREDLTDAIR-----LPPSGRK- 390		
DB	1261	YRCLYDGF	M---ASEDNKTCVDVNECD	LNPNICLSGTCENTKGSFICHCDMGYSKKG 1316		
QY	391	---KDCPPCNP	GFYN-----NGSSSCHPCPP	PGTFSDGTK--ECRPCPAGTEPAL 434		
DB	1317	KTGCTDINECE	IGAHCGRHAVCTNTAGSFKC--	SCSPGWIGDGKICTDLDEC	SNHGMC- 1374	
QY	435	GFEYKWNV	LPGNMKTSCFN--VGNSKC--	DGMNGWEVAGDHIQSGAGGSDNDY	LILNLH 490	
DB	1375	-----SQHADCK	NTGMSYRCLCKDGYTG-----	DGFTCTDLDECSENLN 1413		
QY	491	I-----	PGFKPPTSMTGATGSELGRITF	---VFET-----LCSAD 522		
DB	1414	LCNGQCLNAP	GGYRCECDMGFVP--SADGKACED	IDECSLPNICVFGTCHNLPGLFCE 1471		
QY	523	CVLYFMVDIN	RKSTNVVESGGTKEKQAYTH	IEFKNATFT---FT-----WAFQ	RNQG 573	
DB	1472	CEIGY--ELDR	SGGNCTD---VNECLDPTTCISG	NCVNTPGSYTCD	CSPDFELNPTRVG 1525	
QY	574	QDNRRFIND	MVKIY-----SITATNAVD--	GVASSCRACALG-----SEQ 611		
DB	1526	CVDRSGNCYL	NIRPRGDN	GDPTACSNIEIGVGVS	KASCCCSLGKAWGTPCELCP	SVNISEY 1585
QY	612	-----	-----SGSSCV-----	PCPPGHYIEKETNQ 631		
DB	1586	KILCPGGEGR	PNPITVILEDIDECQELP	GLCQGGKCINTFGSFQCR	CP	TGYLYNEDTRV 1645
QY	632	CKE-----	CPPDTYLSHQVYGKEACI	PCGP-----GSKNNQD--	HSVCYSDCFF 674	
DB	1646	CDDVNECET	PGICGPGT---CYNTVGN	YTCI-CPPDYMQVNGN	CMDMRRSICYR-----1697	
QY	675	YHEKENQIL	HYDFSNLSSVGLMNGP	SFTSKGTKYHFFFNISLCG	HIEGKKMALCTNNITD 734	
DB	1698	-----	NYADNQTC	DGELL-----FNMT-----	KKMCCCSYNIRR 1727	

QY 735 FTVKEI----VAGSDDYTNLGVAFVCQSTIIPSESKGFRAALSSQSIILADTFIGVTVET 790  
Db 1728 AWNKCEQCPIPSTDEFATLCG-----SORPGF-----VIDIYTGLPVD- 1766  
QY 791 TLKNINIKEDMFPVPTSQIP-----DVHFFYKSSSTAT----TSCIN 827  
Db 1767 -----IDE-----CREIPGVCENGVCINMVGSRCECPVGFFYNKLLVCEDIDECON 1814  
QY 828 GR-----STAVKMRCNPTKSGAGVIS-----VPSKCPAGTCDGC--TFY 864  
Db 1815 GPVCLRNAECINTAGSYRCD-CKPGYRLTSTGQCNDRNECQEIIPNICSHGQCIDTVGSFY 1873  
QY 865 FL----WESAEACPLCTEHDPEIE-GACKRG 891  
Db 1874 CLCHTGFKTNEDQTMCL--DINECERDACGNG 1903

RESULT 13  
LMB1\_MOUSE STANDARD; PRT; 1786 AA.  
AC P02469;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Laminin beta-1 chain precursor (Laminin B1 chain).  
GN LAMB1-1 OR LAMB-1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87147212; PubMed=3493487;  
RA Sasaki M., Kato S., Kohno K., Martin G.R., Yamada Y.;  
RT "Sequence of the cDNA encoding the laminin B1 chain reveals a  
RT multidomain protein containing cysteine-rich repeats."  
RL Proc. Natl. Acad. Sci. U.S.A. 84:935-939(1987).  
RN [2]  
RP SEQUENCE OF 1292-1786 FROM N.A.  
RX MEDLINE=85051302; PubMed=6209134;  
RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;  
RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of  
RT coiled-coil alpha-helix."  
RL EMBO J. 3:2355-2362(1984).  
RN [3]  
RP SEQUENCE OF 165-172; 539-547 AND 712-719.  
RC STRAIN=BALB/c; TISSUE=Endothelial cells;  
RX MEDLINE=97363207; PubMed=9219532;  
RA Frieser M., Noeckel H., Pausch F., Roeder C., Hahn A., Deutzmann R.,  
RA Sorokin L.M.;  
RT "Cloning of the mouse laminin alpha 4 cDNA. Expression in a subset of  
RT endothelium."  
RL Eur. J. Biochem. 246:727-735(1997).  
CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin  
CC is thought to mediate the attachment, migration, and organization  
CC of cells into tissues during embryonic development by interacting  
CC with other extracellular matrix components.  
CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three  
CC different polypeptide chains (alpha, beta, gamma), which are bound  
CC to each other by disulfide bonds into a cross-shaped molecule  
CC comprising one long and three short arms with globules at each  
CC end.  
CC THE BETA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-  
CC 2 (MEROSIN), AND LAMININ-6 (K-LAMININ).  
CC -!- SUBCELLULAR LOCATION: Extracellular.  
CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR  
CC COMPONENT).  
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
CC -!- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M15525; AAA39407.1; ALT\_INIT.  
DR EMBL; X05212; CAA28839.1; -  
DR PIR; A26413; MMMSB1.  
DR HSSP; P02468; IKLO.  
DR MGD; MGI:96743; Lamb1-1.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001886; LamNT.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR Pfam; PF00053; laminin\_EGF; 13.  
DR Pfam; PF00055; laminin\_Nterm; 1.  
DR PRINTS; PR00011; EGFLAMININ.  
DR ProDom; PD002082; LamNT; 1.  
DR SMART; SM00180; EGF\_Lam; 11.  
DR SMART; SM00136; LamNT; 1.  
DR PROSITE; PS00022; EGF\_1; 9.  
DR PROSITE; PS01186; EGF\_2; 2.  
DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 11.  
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.  
FT SIGNAL 1 21  
FT CHAIN 22 1786 LAMININ BETA-1 CHAIN.  
FT DOMAIN 22 270 LAMININ N-TERMINAL (DOMAIN VI).  
FT DOMAIN 271 334 LAMININ EGF-LIKE 1.  
FT DOMAIN 335 397 LAMININ EGF-LIKE 2.  
FT DOMAIN 398 457 LAMININ EGF-LIKE 3.  
FT DOMAIN 458 509 LAMININ EGF-LIKE 4.  
FT DOMAIN 510 540 LAMININ EGF-LIKE 5 (INCOMPLETE).  
FT DOMAIN 541 772 LAMININ DOMAIN IV.  
FT DOMAIN 773 820 LAMININ EGF-LIKE 6.  
FT DOMAIN 821 866 LAMININ EGF-LIKE 7.  
FT DOMAIN 867 916 LAMININ EGF-LIKE 8.  
FT DOMAIN 917 975 LAMININ EGF-LIKE 9.  
FT DOMAIN 976 1027 LAMININ EGF-LIKE 10.  
FT DOMAIN 1028 1083 LAMININ EGF-LIKE 11.  
FT DOMAIN 1084 1131 LAMININ EGF-LIKE 12.  
FT DOMAIN 1132 1178 LAMININ EGF-LIKE 13.  
FT DOMAIN 1179 1397 DOMAIN II.  
FT DOMAIN 1398 1430 DOMAIN ALPHA.  
FT DOMAIN 1431 1786 COILED COIL (POTENTIAL).  
FT DOMAIN 1216 1315 COILED COIL (POTENTIAL).  
FT DOMAIN 1368 1388 COILED COIL (POTENTIAL).  
FT DOMAIN 1448 1778 BY SIMILARITY.  
FT DISULFID 271 280 BY SIMILARITY.  
FT DISULFID 273 298 BY SIMILARITY.  
FT DISULFID 300 309 BY SIMILARITY.  
FT DISULFID 312 332 BY SIMILARITY.  
FT DISULFID 335 344 BY SIMILARITY.  
FT DISULFID 337 362 BY SIMILARITY.  
FT DISULFID 365 374 BY SIMILARITY.  
FT DISULFID 377 395 BY SIMILARITY.  
FT DISULFID 398 411 BY SIMILARITY.  
FT DISULFID 400 426 BY SIMILARITY.  
FT DISULFID 428 437 BY SIMILARITY.  
FT DISULFID 440 455 BY SIMILARITY.  
FT DISULFID 458 472 BY SIMILARITY.  
FT DISULFID 460 479 BY SIMILARITY.  
FT DISULFID 481 490 BY SIMILARITY.  
FT DISULFID 493 507 BY SIMILARITY.  
FT DISULFID 773 785 BY SIMILARITY.  
FT DISULFID 775 792 BY SIMILARITY.  
FT DISULFID 794 803 BY SIMILARITY.  
FT DISULFID 806 818 BY SIMILARITY.  
FT DISULFID 821 833 BY SIMILARITY.  
FT DISULFID 823 840 BY SIMILARITY.  
FT DISULFID 842 851 BY SIMILARITY.  
FT DISULFID 854 864 BY SIMILARITY.  
FT DISULFID 867 876 BY SIMILARITY.

FT	DISULFID	869	883	BY SIMILARITY.
FT	DISULFID	886	895	BY SIMILARITY.
FT	DISULFID	898	914	BY SIMILARITY.
FT	DISULFID	917	933	BY SIMILARITY.
FT	DISULFID	919	944	BY SIMILARITY.
FT	DISULFID	946	955	BY SIMILARITY.
FT	DISULFID	958	973	BY SIMILARITY.
FT	DISULFID	976	990	BY SIMILARITY.
FT	DISULFID	978	997	BY SIMILARITY.
FT	DISULFID	1000	1009	BY SIMILARITY.
FT	DISULFID	1012	1025	BY SIMILARITY.
FT	DISULFID	1084	1096	BY SIMILARITY.
FT	DISULFID	1086	1103	BY SIMILARITY.
FT	DISULFID	1105	1114	BY SIMILARITY.
FT	DISULFID	1117	1129	BY SIMILARITY.
FT	DISULFID	1132	1144	BY SIMILARITY.
FT	DISULFID	1134	1151	BY SIMILARITY.
FT	DISULFID	1153	1162	BY SIMILARITY.
FT	DISULFID	1165	1176	BY SIMILARITY.
FT	DISULFID	1179	1179	INTERCHAIN (PROBABLE).
FT	DISULFID	1182	1182	INTERCHAIN (PROBABLE).
FT	DISULFID	1785	1785	INTERCHAIN (PROBABLE).
FT	CARBOHYD	120	120	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	356	356	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	519	519	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	677	677	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1041	1041	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1195	1195	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1279	1279	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1336	1336	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1343	1343	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1487	1487	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1533	1533	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1542	1542	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1643	1643	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	1531	1534	SGNA -> MEMP (IN REF. 2).
FT	CONFLICT	1749	1749	D -> N (IN REF. 2).
SQ	SEQUENCE	1786 AA;	196904 MW;	846671B7BF41A474 CRC64;
Query Match 3.2%; Score 172; DB 1; Length 1786;				
Best Local Similarity 19.4%; Pred. No. 0.00017;				
Matches 147; Conservative 61; Mismatches 207; Indels 344; Gaps 41;				
QY	61	COE-KDYHFEYTE-----	CDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFS	106
Db	374	CEQCKPFYFQHPERDIRDPNLCBPCTCDPAGSE-----	NGGI-CDGYTD-----	FS 418
QY	107	CA--SGE--YLEMKNQVCSKCGEGTYSLSGGIKFDEWDELPAFGFSNIA-TFMDTVVG--	158	
Db	419	VGLIAGQCRCKLHVEGERCDVCKEGFYDLA-----	EDPYGCKSCACNPLGTIPGSGN	470
QY	159	PSDSRP-----	DGC-----	NNS-----S 171
Db	471	PCDSEGYCYCKRLVTGQRCDQCLPQHWGLSNDLDGCRPCDCDLGGALNNSCSEDSGQCS	530	
QY	172	WIP-----RGNYESNRDDCTVS-LIYAVHLKK--SGYVFEYQYVDNNI-----	213	
Db	531	CLPHMTGRQCNEVESGYFTTLQHYIYEAEANLPGVVVVERQYIQDRIPSWTGPGEFVR	590	
QY	214	-----PFEFFIQN-DQCQEMDTTDDKWVKLTONGEWGSHVMLKSGTNI LYWRTTGILMG	267	
Db	591	VPEGAYLEFFIDNIPYSMEYEILIRYQLPDHWE-----	625	
QY	268	SKAVKPVLVKNITIE--GVAYTSECFPCPKPG---TFSNKPFSFNCQVCPRNTYSEKGA	320	
Db	626	-KAV-----ITVQRPKIPASSRCGNTVPDDDNQVVSLSPGS-RYVVLP RPVCFEKGM	676	
QY	321	KECIRCKDDSQFSGSSECTE-----	RPPCTTKDYQFIHTPCDEE-----	GKTQI 364
Db	677	NYTVRL-ELPQYTASGSDVESPYTFIDSLVLMPYCKSLDIFTVGGSGDGEVNTSAWETFQ	735	
QY	365	MYKWIE--PKICREDLTDAIR-----	LPPSGEKKDCPP-----	CNPGFYNNGSS 406

Db	736	RYRCLENSRSYVKTPMTDVCRNIIIFSISALIHQTGLACECDPQGSLSVCDP----	NGGQC 799
QY	407	SCHP-----CPPGTESDGTKECRPCPCAGTEPALGFYKWNVLPGNMKTSCFNVGN	457
Db	793	QCRPNVVGRTCNRCAPGTGFGPNGCKPC-----	DCHLQGS 828
QY	458	SK--CDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMGTGATGSELGRITFVF	515
Db	829	ASAFCDAITG---QCHCFQGIYARQCDCRC-----	LPGY----- 858
QY	516	ETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTTWFQFQRTNQGD	575
Db	859	-----	WGFPSCQPCQC 869
QY	576	NRRFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGHY---IEKETNOC	632
Db	870	NGHALD-----	CDTVTGECLSCQ--DYTTGHNCRCLAGYYGDPPIIGSGDHC 914
QY	633	KECPPDTYLSIHQVYGKEACIPCGPGSKNNQDHSVCYSD	671
Db	915	RPCP-----	CPDGPDS-GRQFARSCYQD 936
RESULT 14			
NTCL_BRARE . STANDARD: PRT: 2437 AA.			
ID	NTCL_BRARE		
AC	P46530;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Neurogenic locus notch homolog protein 1 precursor.		
GN	NOTCH1A OR NOTCH.		
OS	Brachydanio rerio (Zebrafish) (Danio rerio).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;		
OC	Cyprinidae; Danio.		
OX	NCBI_TaxID=7955;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Embryo;		
RX	MEDLINE=94128602; Pubmed=8297791;		
RA	Bierkamp C., Campos-Ortega J.A.;		
RT	"A zebrafish homologue of the Drosophila neurogenic gene Notch and its pattern of transcription during early embryogenesis.";		
RL	Mech. Dev. 43:87-100(1993).		
CC	-!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.		
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-!- DEVELOPMENTAL STAGE: Expressed in all cells in pregastrulation stages. During gastrulation is differentially expressed, accumulating predominantly in the prechordal mesoderm and notochord. At the end of gastrulation, expressed along the anterior-posterior axis including the developing neural plate and differentiating mesoderm. Also present in the developing brain and head regions.		
CC	-!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.		
CC	-!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.		
CC	-!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.		
CC	-!- SIMILARITY: CONTAINS 6 ANK REPEATS.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
CC	EMBL; X69088; CAA48831.1; -		
DR	HSSP; P00740; 1EDM.		
DR	ZFIN; ZDB-GENE-990415-173; notchla.		
DR	InterPro; IPR002110; ANK.		





FT	DISULFID	909	920	BY SIMILARITY.
FT	DISULFID	914	929	BY SIMILARITY.
FT	DISULFID	931	940	BY SIMILARITY.
FT	DISULFID	947	958	BY SIMILARITY.
FT	DISULFID	952	967	BY SIMILARITY.
FT	DISULFID	969	978	BY SIMILARITY.
Query Match 3.2%; Score 170.5; DB 1; Length 2437;				
Best Local Similarity 19.7%; Pred. No. 0.00033;				
Matches 233; Conservative 102; Mismatches 374; Indels 473; Gaps 71;				
QY	42	QAAWAGDL	-----PSSSRPLPPCQEKDYHFEYTECDSSGRWRVAIP	-----NS 87
Db	128	QPGWSGKTCQLADPCASN	-----PCANGG-----QCSAFESHYICTPPNFHGQTCRQD	177
QY	88	AVDCSGLPDVR-GKEC	-----TFSCA-----SGEYLEMKNQVC--SKCGEGTYSLGSGI	134
Db	178	VNECAVSPSPCRNGGTCTINEVGSYLRCRCPPEYTGPHCQRLYQCLPSPCRSG	-----	229
QY	135	KFDEWDELPAFGFSNIATFMDTVVGPDSRDPDGCNNSWIP	-----RGNYIESNRDDCTVSL--	190
Db	230	-----GTCVQTSDT	-----THTCSCLPGETGTQCEHNVDDCTQHACE	266
QY	191	-----IYAVHLKSGYVFFEQYVDNNIFFEFTIQNDQCQEMDT	-----TTDKWVKL	237
Db	267	NGGPCIDGINTYNCHCDK	-----HWTGQYCTEDV-DECELSPNACQNGGTCHNTIGGFHCV	321
QY	238	TDNGEWG	-----SHSVMLKSGTNILYWR	265
Db	322	CVNGWTGDDCSENIDDCASASHGATCHDRVASFFCECPHGRTGLLCHLDDACISNPCQ	381	
QY	266	MGSKA-VKPVLVKNITIEGVAYT	-----SECF-----PCK-PGTFSNKPGSFNCQVC-	310
Db	382	KGSNCDTNVPSGKAICTCPPGYTGSAQNQDIDECSLGANPCHEHGRCLNKGSTQCKCLQ	441	
QY	311	-----PRNTYSEKGAKECIR--CKDDS	-----QFSG-----SSECTERPP	343
Db	442	GYEGPR--CEMDVNECKSNPCQNDATCLDQIGGFHCICMPGYEGVFCQINSDDCASQPC	498	
QY	344	CTTKDYFQI--HTPCDEEGKTQIMYKWIPIKICREDLTDAIRLP	-----PSGEKKDC	393
Db	499	LNGKCIDKINSFHCECP	-----KFGSGLCQVDVDECASTPCKNGAKCTDGPKNKYT	549
QY	394	PPCNPGR	-----YNNGSSSCHPCPPGTFSDG-----TKECRP	425
Db	550	CECTPGFSGIHCELDINECASS--PCHYGVCRDGVASFCTCDRCRPGYTGRLCETNINECLS	607	
QY	426	-----CPAGTEPALGFYKWNVLPGNMKTSC-FNVGN	-----	457
Db	608	QPCRNGGTQDRENAYICTCPKGT	-----TGVNCEINIDCKRKPCDY	650
QY	458	SKC-DGMNGWEVAGDHIQSGAGGSDN-DYLILN	-----LHIPGFKPPT	498
Db	651	GKCIDKINGYECVCEPGYSGSMCNINIDDCALNPCHNGGTCTIDGVNSFTCLCPDGFDRAT	710	
QY	499	SMTGAT	-----GSELGRITTFVFETLCSA-----DCVLYFMVDINRKSTNVVESWGG	544
Db	711	CLSQHNECSSNPCIHGSCLDQIN-SYRCVCEAGWMGRNC	-----DININECLSNPCVN-GG	764
QY	545	TKEKQAYTHIIFKNATFTTTFWAFQRTNOGQDNRFRINDMVKIYSITATNAVGVASSCRA	604	
Db	765	TCKDMTSGYLCTCRAGFS	-----GPNQCMNINECASNPCLNQGSCIDDDVAGFKCN	814
QY	605	CALGSEQSGSSC	-----VPCPPGHYIEKETNOCKE-----CPPDTYLSIHQVYGKE	650
Db	815	CML--PYTGEVCENVLAPCSRPC--KNGGVCRESEDFQSFSCNCPAGWQGTCEVDINE	870	
QY	651	ACI--PCGPGSKNNQDHSVCSYDCFFHYHEKENQILHYDFSLSVSGSLMNGPSTSKGTK	708	
Db	871	-CVRNPCTNG	-----GVC-----ENLRGGEQCRCPNPGFTG	899
QY	709	YFHFFNISLCGHEGKKMALCTNNITDFTVKEIVAGS--DDYTNLVGAFVQCSTIIPSESK	766	

Db	900	-----ALCENDIDDCEPNPCSNNGVCQDRVN	-----GFVCVCL	-----A 933
QY	767	GFERAALSSQSIIADTFIGTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSSTATTS	CI 826	
Db	934	GFRGERCAEDI	-----DECVSAPC	-----RNGNCITDCV 962
QY	827	NGRSTAVKMRNPTKSGAGV	---ISVP---	SKC-PAGTC-DGCTFYFLWESAECPLCT 877
Db	963	NSYT	-----CSCPAGFSGINCEINTPDCTESSCFNGGTCVDGSISSF	-----SCVCLPGFT 1012
QY	878	---EHDHFEIE	-----GACKRGFQ	-----ETLYVWNEPKWCJKIGISL 911
Db	1013	GNYCQHDVNECDSPRCQNGGSCQDGYGTYKCTCPHGTYTGLNCQSLVRWCDS	SPCKNGGSC 1072	
QY	912	PEKCLA-TCETVDFWL	-----KVGAGVGAFITAVLLVALTC 945	
Db	1073	WQOGASFCTCQCASGWTGIYCDVPSVSCVAAARQGVSAVLC	1114	
RESULT 15				
LML1_CAEEL STANDARD; PRT; 1557 AA.				
AC	Q18823;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	Laminin-like protein C54D1.5 precursor.			
GN	C54D1.5.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;			
OC	Rhabditidae; Peloderinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Bristol N2;			
RA	Minx P.;			
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.			
CC	- - SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).			
CC	- - SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.			
CC	- - SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@lsb-sib.ch">license@lsb-sib.ch</a> ).			
CC	-----			
DR	EMBL; U46673; AAC48152.1; -.			
DR	HSSP; P02468; 1TLE.			
DR	WormPep; C54D1.5; CE06981.			
DR	InterPro; IPR000561; EGF-like.			
DR	InterPro; IPR001886; LamNT.			
DR	InterPro; IPR000034; Laminin_B.			
DR	InterPro; IPR002049; Laminin_EGF.			
DR	Pfam; PF00052; laminin_B; 1.			
DR	Pfam; PF00053; laminin_EGF; 10.			
DR	Pfam; PF00055; laminin_Nterm; 1.			
DR	PRINTS; PR00011; EGFLAMININ.			
DR	ProDom; PD002082; LamNT; 1.			
DR	ProDom; PD003031; Laminin_B; 1.			
DR	SMART; SM00180; EGF_Lam; 9.			
DR	SMART; SM00001; EGF_like; 2.			
DR	SMART; SM00281; LamB; 1.			
DR	SMART; SM00136; LamNT; 1.			
DR	PROSITE; PS00022; EGF_1; 8.			
DR	PROSITE; PS01186; EGF_2; 1.			
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 11.			
KW	Hypothetical protein; Laminin EGF-like domain; Signal; Repeat.			
FT	SIGNAL 1 ?			
FT	CHAIN ? 1557			
FT	DOMAIN ? 271			

FT	DOMAIN	272	331	LAMININ EGF-LIKE 1.
FT	DOMAIN	332	387	LAMININ EGF-LIKE 2.
FT	DOMAIN	388	434	LAMININ EGF-LIKE 3.
FT	DOMAIN	435	487	LAMININ EGF-LIKE 4.
FT	DOMAIN	488	497	LAMININ EGF-LIKE 5 (N-TERMINAL).
FT	DOMAIN	498	688	LAMININ DOMAIN IV.
FT	DOMAIN	687	732	LAMININ EGF-LIKE 5 (C-TERMINAL).
FT	DOMAIN	737	769	LAMININ EGF-LIKE 6 (INCOMPLETE).
FT	DOMAIN	770	819	LAMININ EGF-LIKE 7.
FT	DOMAIN	820	874	LAMININ EGF-LIKE 8.
FT	DOMAIN	875	930	LAMININ EGF-LIKE 9.
FT	DOMAIN	931	978	LAMININ EGF-LIKE 10.
FT	DOMAIN	979	1025	LAMININ EGF-LIKE 11.
FT	DISULFID	272	281	BY SIMILARITY.
FT	DISULFID	274	295	BY SIMILARITY.
FT	DISULFID	297	306	BY SIMILARITY.
FT	DISULFID	309	329	BY SIMILARITY.
FT	DISULFID	332	341	BY SIMILARITY.
FT	DISULFID	334	357	BY SIMILARITY.
FT	DISULFID	360	369	BY SIMILARITY.
FT	DISULFID	372	385	BY SIMILARITY.
FT	DISULFID	388	400	BY SIMILARITY.
FT	DISULFID	390	406	BY SIMILARITY.
FT	DISULFID	408	417	BY SIMILARITY.
FT	DISULFID	420	432	BY SIMILARITY.
FT	DISULFID	435	449	BY SIMILARITY.
FT	DISULFID	437	456	BY SIMILARITY.
FT	DISULFID	458	467	BY SIMILARITY.
FT	DISULFID	470	485	BY SIMILARITY.
FT	DISULFID	770	778	BY SIMILARITY.
FT	DISULFID	772	788	BY SIMILARITY.
FT	DISULFID	791	800	BY SIMILARITY.
FT	DISULFID	803	817	BY SIMILARITY.
FT	DISULFID	820	834	BY SIMILARITY.
FT	DISULFID	822	841	BY SIMILARITY.
FT	DISULFID	844	853	BY SIMILARITY.
FT	DISULFID	856	872	BY SIMILARITY.
FT	DISULFID	875	894	BY SIMILARITY.
FT	DISULFID	877	901	BY SIMILARITY.
FT	DISULFID	903	912	BY SIMILARITY.
FT	DISULFID	915	928	BY SIMILARITY.
FT	DISULFID	931	943	BY SIMILARITY.
FT	DISULFID	933	950	BY SIMILARITY.
FT	DISULFID	952	961	BY SIMILARITY.
FT	DISULFID	964	976	BY SIMILARITY.
FT	DISULFID	979	991	BY SIMILARITY.
FT	DISULFID	981	998	BY SIMILARITY.
FT	DISULFID	1000	1009	BY SIMILARITY.
FT	DISULFID	1012	1023	BY SIMILARITY.
FT	CARBOHYD	101	101	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	121	121	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	333	333	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	507	507	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	643	643	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	725	725	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	921	921	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1062	1062	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1168	1168	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1211	1211	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1244	1244	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1321	1321	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1437	1437	N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE	1557 AA;	17273 MW;	CAF0B51F8D5E8D2F CRC64;

Query Match 3.2%; Score 170; DB 1; Length 1557;  
Best Local Similarity 18.1%; Pred. No. 0.0002;  
Matches 173; Conservative 86; Mismatches 292; Indels 404; Gaps 49;

QY	42	COAAWAGDLPSSSRPLPQCQEKDYHFEYTECDSSGSRWRVAIPN-----SAVDCSGLP	95
Db	406	CQCK-----PGVTGRFCDCQCLDGGFYDFSTNGCKNCGETSGSLNNQPRCDSSSGSCSKL	460
QY	96	DPVRGECTFSCASGEY-LEMKNQV-CSKC-----GEGTSLGSGIKFDEWDEL	142

Db	461	N-VEGRQCD-KCKPGYFDLSTENQFGCTPCFCFHSSICNTADGYFAMNVSSVFDQDKQK	518
QY	143	PAGFSNIA-----TFMDTVVGPSDSRPDGCNNSS--WIPRGNVIESNRDDCTVSLIYAV	194
Db	519	WAGONRIGLQDTQWAECLKAVAVSDT-----DNSPVYFVAPEQFLGDQRSSYNQDLVFTL	573
QY	195	HLKSGYVFFEYQYVDNNIFFEFFFQNDQCQEMDTTDDKWKLTDNGE-----	242
Db	574	KVAK-----HVTNQDVKDIIIVGADRQELSTS-----ITAQGNPFPPTTEAQT YRF	618
QY	243	-----WGSHSVMLKSGTNILYWRTT-----GILMGSKAVKPVLVK-----	277
Db	619	RVHADPYFGW-----YPRINELDFIGILSNITAIK---IRGTYSYKDIGYL	661
QY	278	---NITIEGVAYTS-----ECFPCKPGTFSNK-----PGSFN-----	306
Db	662	SNVNLGTAGVAPSAANPKQATWIEHCECLPGFVGQFCESCESGFRRET KFGGPFNHCICK	721
QY	307	-----CQ-----VCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQI	352
Db	722	DCHNHSNCEAESGSCICEHNTAGDT-CERCARGYYGDALQGTEDCQKCPCPNDGPCIL	780
QY	353	HTPCDEEGKTQIMYKWIPEKICREDLTDAIRLPPSGEKKDCPPCNPFGFYNNNGSSSCHPCP	412
Db	781	HADGD-----VICTE-----CPNGYTGRRCDECDGDFGNP-----	811
QY	413	PGTFSDBGTK--ECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAG	470
Db	812	----KDGTECVEC-ACSGNTDPN-----SIGN--CDKITG-----	839
QY	471	DHIQSGAGGSDNDYLILNLHIPGPKPPTSMTGATGSELGRITVFETLCSADCVL YFMVD	530
Db	840	-----ECKKCIFNTH--GFNCENCKPGYWGDA-----	865
QY	531	INRKSTNVVESWGGTKEKQAYTHIFKNATFTTWAQRTNQODNRRFRINDMVKIYSIT	590
Db	866	-----IEPKGNCQSCGF-----AAGTRRPND-----YTLL	892
QY	591	ATNAVVG-----VASSCRACALG--SEQSGSCVPC-----PPGHYIEKETNOCKE	634
Db	893	ECNQDGGQCDCLPNVIGIQDQCAHGFYNITSLGLCQECNCDDL GSEGN TCDVNTGQC-Q	951
QY	635	CPP-----DTYLSIHQVYGKEACIPCG---PGSKN-----NQDHSVVCYSDCFFYHEKE	679
Db	952	CKPGVTGQRCDRCADYHGFSGANGCQPCDCÉYIGSENOQQCDVNSGQCLC-----KE	1002
QY	680	NQILHYDFSNLSSVGLMNGPSFTSKGTKYFHFFNISLCGHEGKKMALCTNNITDFTVKE	739
Db	1003	NV-----EGRRCDOCAEN-----	1015
QY	740	IVAGSDDYTNLVGAFVCQS--TIIPSESKGFRAALSSQSIIILADTFIGVTVETTLKNINI	797
Db	1016	-----RYGITQGCLPCDDCYTLIQSRVNVFREKVK S-----LDNTLQEIEN---	1057
QY	798	KEDMFPVPPTSQIPDVHFFYKSSSTATTSCTINGRSTAVKMR CNPTKSGAGVISVPSK	852
Db	1058	-----PAPVNDTK----FDEKVKETSRAASEVWEAVKQK----TKEGGGTIKTKSK	1100

Search completed: May 12, 2003, 13:19:58  
Job time : 37.3568 secs





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.  
OM protein - protein search, using sw model  
Run on: May 12, 2003, 12:50:23 ; Search time 41.1332 Seconds  
(without alignments)  
4823.927 Million cell updates/sec  
Title: US-10-073-333A-2  
Perfect score: 5357  
Sequence: 1 MLFRARGPVRGRGWGRPAEA.....TCYFWKKNQKKKKTILNLFN 963  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length	ID	Description
1	2624	49.0	493	Q96DP2	Q96dp2 homo sapien
2	1557	29.1	580	Q9P2M2	Q9p2m2 homo sapien
3	1225.5	22.9	1019	Q9NA40	Q9na40 caenorhabdi
4	661.5	12.3	300	11 Q8R215	Q8r215 mus musculu
5	222.5	4.2	1074	5 Q964D1	Q964d1 entamoeba h
6	211.5	3.9	1101	5 Q964D2	Q964d2 entamoeba h
7	207.5	3.9	704	5 Q9U048	Q9u048 giardia lam
8	203	3.8	709	5 Q9XTJ7	Q9xtj7 giardia lam
9	201	3.8	709	5 Q97444	Q97444 giardia lam
10	199.5	3.7	667	5 Q95WU1	Q95wu1 giardia lam
11	197.5	3.7	3087	5 Q8WRD2	Q8wrd2 plasmodium
12	193.5	3.6	5374	11 Q99ND0	Q99nd0 mus musculu
13	192.5	3.6	719	5 Q9U021	Q9u021 giardia lam
14	192.5	3.6	2759	5 O45614	O45614 caenorhabdi
15	192.5	3.6	3102	5 Q9TZR4	Q9tzt4 caenorhabdi
16	191	3.6	1289	5 Q8SSS3	Q8sss3 dictyosteli

17	188.5	3.5	769	5 Q24971	Q24971 giardia lam
18	188.5	3.5	999	4 Q9NQ36	Q9nq36 homo sapien
19	187	3.5	1723	5 Q8WRD0	Q8wrd0 plasmodium
20	186	3.5	1007	13 Q90ZN3	Q90zn3 gallus gall
21	185.5	3.5	719	5 Q9U019	Q9u019 giardia lam
22	185.5	3.5	2534	5 Q8WRD1	Q8wrd1 plasmodium
23	184	3.4	1557	5 Q8WPK9	Q8wpk9 oikopleura
24	183	3.4	1372	5 P91526	P91526 caenorhabdi
25	182	3.4	2931	5 Q9W2C6	Q9w2c6 drosophila
26	181.5	3.4	1551	5 Q9NGV4	Q9ngv4 drosophila
27	181.5	3.4	3396	5 Q9VM55	Q9vm55 drosophila
28	181	3.4	2189	5 Q9BI05	Q9bi05 eimeria ten
29	179.5	3.4	718	5 Q9BI07	Q9bi07 entamoeba h
30	179.5	3.4	3567	11 Q9ES77	Q9es77 mus musculu
31	178.5	3.3	3704	5 P91904	P91904 caenorhabdi
32	175	3.3	2906	11 Q9WUH9	Q9wuh9 rattus norv
33	174.5	3.3	548	5 Q9GQ45	Q9gq45 giardia lam
34	174.5	3.3	597	5 Q07317	Q07317 giardia int
35	173.5	3.2	468	5 Q9BMH3	Q9bmh3 ichthyophth
36	173	3.2	667	5 Q9XTK3	Q9xtk3 giardia lam
37	173	3.2	1045	5 Q8T3A6	Q8t3a6 caenorhabdi
38	173	3.2	1070	5 Q8T3A7	Q8t3a7 caenorhabdi
39	173	3.2	1111	5 Q9XWD6	Q9xwd6 caenorhabdi
40	172.5	3.2	3857	11 Q88840	Q88840 mus musculu
41	171.5	3.2	2809	4 Q96JP8	Q96jp8 homo sapien
42	171	3.2	1316	4 Q96JU7	Q96ju7 homo sapien
43	169.5	3.2	739	5 Q9GS24	Q9gs24 giardia lam
44	168	3.1	594	5 Q24970	Q24970 giardia lam
45	167.5	3.1	1664	5 Q9TVQ2	Q9tvq2 caenorhabdi

ALIGNMENTS

RESULT 1  
Q96DP2 ID Q96DP2 PRELIMINARY; PRT; 493 AA.  
AC Q96DP2;  
DT 01-DEC-2001 (TremBLrel. 19, Created)  
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)  
DE CDNA FLJ31340 fis, clone MESAN1000035, weakly similar to major  
DE surface-labeled trophozoite antigen precursor.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,  
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,  
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,  
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,  
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,  
RA Isogai T.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK055902; BAB71041.1;  
DR InterPro; IPR001881; EGF\_CA.  
DR SMART; SM00179; EGF\_CA; 1.  
SQ SEQUENCE 493 AA; 54585 MW; 0AEBB2C19709B211 CRC64;

Query Match 49.0%; Score 2624; DB 4; Length 493;  
Best Local Similarity 97.3%; Pred. No. 3.2e-215;  
Matches 470; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

QY	244	GSHSVMLKSGTNILYWR	TTGILMGSKAVKPVLV	KNITIEGVAYTSECF	PKPGTFSNKP	303
Db	4	GNYQVMLKSGTNILYWR	TTGILMGSKAVKPVLV	KNITIEGVAYTSECF	PKPGTFSNKP	63
QY	304	SFNCQVCPRNTYSEKGA	KECIRCKDDSQFS--	GSSECTERPPCTTKDY	FQIHTPCDEEG	361

Db 64 SFNCQVCPNTYSEKAGEKIRCKDDSQFSEGSSECTERPPCTTKDYFQIHPTPCDEEGK 123  
QY 362 TQIMYKWIPIKICREDLTDAIRLPPSGEKKDCPPCPNGFYNGSSSCHPCPPGTFSDGTK 421  
Db 124 TQIMYKWIPIKICREDLTDAIRLPPSGEKKDCPPCPNGFYNGSSSCHPCPPGTFSDGTK 183  
QY 422 ECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSD 481  
Db 184 ECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSD 243  
QY 482 NDYLILNLHIPGFKPPTSMTGATGSELGRITVFVETLCSADCVLVFMVDINRKSTNVVES 541  
Db 244 NDYLILNLHIPGFKPPTSMTGATGSELGRITVFVETLCSADCVLVFMVDINRKSTNVVES 303  
QY 542 WGGTKEKQAYTHIIFKNATFTFTWAFQRTNQGDNRFRINDMVKIYSITATNAVDGVASS 601  
Db 304 WGGTKEKQAYTHIIFKNATFTFTWAFQRTNQGDNRFRINDMVKIYSITATNAVDGVASS 363  
QY 602 CRACALGSEQSGSSCVPCPPGHYIEKETNQCKECPDPTVLSIHQVYGKEACIPCGPGSKN 661  
Db 364 CRACALGSEQSGSSCVPCPPGHYIEKETNQCKECPDPTVLSIHQVYGKEACIPCGPGSKN 423  
QY 662 NQDHSVCYSDCFFYHEKENQILHYDFSNNLSSVSGSLMNGPSTSGTKYFHFHNISLCGHE 721  
Db 424 NQDHSVCYSDCFFYPEKENQSLHYDFSNNLSSVSGSLMNGPSTSGTKYFHFHNISYVGMR 483  
QY 722 GKK 724  
Db 484 GRR 486

RESULT 2  
Q9P2M2 PRELIMINARY: PRT; 580 AA.  
AC Q9P2M2;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE KIAA1324 protein (Fragment).  
GN KIAA1324.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.

RC TISSUE=BRAIN;  
RX MEDLINE=20181126; PubMed=10718198;  
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes.XVI.  
RT The complete sequences of 150 new cDNA clones from brain which code  
RT for large proteins in vitro."  
RL DNA Res. 7:65-73(2000).  
DR EMBL; AB037745; BAA92562.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 580 AA; 63365 MW; 5E63C19265EC8E5C CRC64;

Query Match 29.1%; Score 1557; DB 4; Length 580;  
Best Local Similarity 52.1%; Pred. No. 4.4e-124;  
Matches 274; Conservative 93; Mismatches 151; Indels 8; Gaps 3;

QY 440 WNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTS 499  
Db 1 WNTLPTNMETVLGSGINFEYKGMTGWEVAGDHIYTAAGASDNDFMILTLPVPGFRPPQS 60  
QY 500 -MTGATGSELGRITVFETLCSADCVLVFMVDINRKSTNVVESWGGTKEKQAYTHIIFKN 558  
Db 61 VMADTENKEVARITVFETLCSVNCELYFMVGNSRNTNPVETWKGSKGQSYTHIIEEN 120  
QY 559 ATFTFTWAFQRTNQGDNRFRINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVP 618  
Db 121 TTTSTWAFQRTTFHEASRKYTNDVAKIYSINVTNVMNGVASYCRPCALEASDVGSCTS 180

QY 619 CPPGHYIEKETNQCKECPDPTVLSIHQVYGKEACIPCGPGSKNNQDHSVCYSDCFFYHEK 678  
Db 181 CPAGYYIDRDSGTCHSCPNTILKAHQPYGVQACVPCPGGTKNKIHSLCYNDCTFSRNT 240  
QY 679 ENQILHYDFSNNLSSVSGSLMNGPSTSGTKYFHFHNISLCGHEGKKMALCTNNITDFTVK 738  
Db 241 PRTENYNFSALANTVTLAGGPFSTSGTKYFHFHFTLSLCGNQGRKMSVCTDNTDLRIP 300  
QY 739 EIVAGSDDYTNLVGAFCVQSTIIPSESKGFRAALSQSIIADTFIGVTVETTLKNINIK 798  
Db 301 E---GESGFSKSIYVCOAVIIPPEVTGYKAGVSQPVSLADRLIGVTTDMTLDGITSP 357  
QY 799 EDMFPVPTSQIPDVHFFYKSTATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGTC 858  
Db 358 AELFHLESGLIPDVIFFYRSNDVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCSDGTC 417  
QY 859 DGCTFYFLWESAEACPLCTEHDHFHEIEGACKRGFQETLYVWNEPKWCIKGISLPEKKLAT 918  
Db 418 DGCNHFELWESAAACPLCSVADYHAIIVSSCVAGIQKTTYVWREPKLCSGGISLPEQRTI 477  
QY 919 CERVDFWLKVGAGVGAFTAVLLVALTCYFWKKNQKKK---KTILN 960  
Db 478 KTIIDFWLKVGISAGTCTAILLTVLTCYFWKKNQKLEYKSKLVMN 523

RESULT 3  
Q9NA40 PRELIMINARY: PRT; 1019 AA.  
AC Q9NA40;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE Y73F8A.5 protein.  
GN Y73F8A.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Matthews L.;  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;

RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
RT investigating biology."  
RL Science 282:2012-2018(1998).  
DR EMBL; AL132862; CAB70224.1; -.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; UNKNOWN\_1.  
SQ SEQUENCE 1019 AA; 112378 MW; 6EB4094722B707E7 CRC64;

Query Match 22.9%; Score 1225.5; DB 5; Length 1019;  
Best Local Similarity 31.5%; Pred. No. 1.9e-95;  
Matches 298; Conservative 155; Mismatches 371; Indels 123; Gaps 36;

QY 61 CQEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120  
Db 25 CTQDQDYFTYTNCDENGERWRVAVPRGMQCSNLPTPRRLNCSFSCEPGHYLDLDSQHC 84  
QY 121 SKCGEGTYSLGSGIKFDEWDELPAFNSIATFMDTVVGPSPDSRPDG-----C-N 168  
Db 85 RPCNPGFFSLGGGIRYEEFVTLPSGFS-----VDNMDSNPDAQFSNRQSQVVECPK 135  
QY 169 NSSWIPRGNYIESNRDDCTVSLIYAVHLKKSGYVEFEYQYVDNN--IFFEFFIONDQCQE 226  
Db 136 EAGWVVKDGLIYIPTPCVSKLSFSANLVRPGSVETTYRMPRNRNALSQMVDIRNEQCQS 195  
QY 227 M-DTTTTDKWVKLT-----DNCEWGSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVK 277



Db 196 YNDVAKSMFLKYTKKEDEEKNGDWRKRRRIELKSGANVISWIIQNNMGYQASNQPIHID 255  
QY 278 NITIEGVAYTSECFCKPGTFSNKPGEFNCQVCPRNTYSEKGAKECIRCKDDSQFSG--S 335  
Db 256 RIDVLGLAFTRQCTACPPGT-SSPGSAECIPCSPGSSFKSGSQCGRC-PESQYSGFKS 313  
QY 336 SECTERPPCTKDYFQIHTPCDEEGKTQIMYKWIETKICREDLTDAILPPSGEKKDCPP 395  
Db 314 EKCIDRPPCRVSDYYPVREPC-TNGSSRAVYKVLPSICRDDLP SATKLPPPTPWKTCPK 372  
QY 396 CNPGFYNNSSSCHPCPPGTFSDGTEKCRPCPAGTEPALGFEFYKWNVLPGNMKTSCFNV 455  
Db 373 CNPGMEKNKLGVCFECKDYFSDG-NSCSRCPVDTPVNYGLQYQNWEVLPKLSRCEYI 431  
QY 456 GN---SKCDGMNGWEVAGDHIOQGAG---GSDNDYLILNLHIPGF-----KPPTSMTGA 503  
Db 432 SEDVATACNIGDAWIPSGDSLISAPLSLELGIAFE-LILSID-EGEWNPLAPKPSKTMK-- 487  
QY 504 TGSELGRITFVFETLCSAD-CVLYFMVDIN-----RKS-TNVVESWGGTKEQAYTHII 555  
Db 488 --VPVAQVTIVFETSCADESCALYFIEDMSAGIKGQRESEFYHFLAAFNNGSNKRVSHTV 545  
QY 556 FKNATFTTFAFQRT--NQGDNRFRFINDVMVKIYSITATN-----AVDGVASSCRACALG 608  
Db 546 TKNTPAREMVAFLRSGVSSGDDK---ISDETRIYAINTNVGHRGQGGGASQCLTCP-- 600  
QY 609 SEQSGSSVCPPPGHYIEKETNOCKECPDPTYLSI-HQVYGKEACIPCGPGSKNNQDHSV 667  
Db 601 HTAGGETCVCPCPAGNYMHEVTKLVCSPPLNTIINVTSSRVGVKSCVPCGQGLTSNDGVSC 660  
QY 668 CYSDCFFYHE----KENQILHYDFS-----NLSSVGLMNGPFSFTSGTKYFHFENI 715  
Db 661 TAMGKIQLNQIGGKGNDSSTYDFSPFVGRSWNISGV-----RVFSREGAAYYHFFSV 714  
QY 716 SLCGHEGKKNALCTNNITDFTVKEIVAGSDDYTNLVGAFVCQSTIIPSESKGFRAALSSQ 775  
Db 715 SLFPPNPK-----CQEQFDNFMIGIL--DQDKENVEG-LACRVLTALTPSSNRSKTAYVT 767  
QY 776 SIILADTFIGVTVETTLKNINIKEDMEFPVP-----TSQIPDVHFFYKS-STATTSCINGRS 830  
Db 768 PLLLAGRLDISLNRKKGNSLSDEVLEYSHDNTSHPLDVFFFEFVPVSTISPACPNGNQ 827  
QY 831 TAVKMRNPTKSGAGVISVPSKCPAGTCGCTGYFLWESAEACPLCTEHDHFHEIEGACKR 890  
Db 828 LVVVARCVPTKKQM-EMRLPHNCPDGTCDGCLFVIIMETAQACPVCESNDYETINGECLN 886  
QY 891 GFOETLYVWNEPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAFTA 937  
Db 887 G-KQTIHSIPKKHCVITGAASQSKEVA-----CSAFTA 918

RESULT 4  
Q8R215  
ID Q8R215 PRELIMINARY; PRT; 300 AA.  
AC Q8R215;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical 32.6 kDa protein (fragment).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=COLON;  
RA Strausberg R.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC022655; AAH22655.1; -.  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 300 AA; 32642 MW; 6455109C054B6CB4 CRC64;

Query Match 12.3%; Score 661.5; DB 11; Length 300;  
Best Local Similarity 49.4%; Pred. No. 4.5e-48;  
Matches 117; Conservative 43; Mismatches 70; Indels 7; Gaps 2;  
QY 728 CTNNITDFTVKEIVAGSDDYTNLVGAFVCQSTIIPSESKGFRAALSSQSIILADTFIGVT 787  
Db 1 CTDNVTDLRIPDGEAG--FSKSVTAYVCQVVIIPSEVMGYKAGVSSQPVSLADRLVGVS 57  
QY 788 VETTLKNINIKEDMEFPVPTSQIPDVHFFYKSSATTSCINGRSTAVKMRNPTKSGAGVI 847  
Db 58 TDMTLEGIVSPVELFHPETSGIPDIVFFFRSNDVTQSCSSGRSTIIRLCNPMKAAPGTL 117  
QY 848 SVPSKCPAGTCGCTFYFLWESAEACPLCTEHDHFHEIEGACKRGFQETLYVWNEPKWCIK 907  
Db 118 RLPSMCSGDTCDGCNHFHFWESVAACPLCSASDYHTFVSSCVAGIQKTTYMWREPKLCSG 177  
QY 908 GISLPEKKLATCETVDFWLKVGAGVGAFTAVALLVALTCTYFKNQKKK----KTILN 960  
Db 178 GISLPEQRVTICKTIDFWLKVGISAGTCTAILLTVLTCTYFWRKNQKLEYKYSKLVMN 234  
RESULT 5  
Q964D1  
ID Q964D1 PRELIMINARY; PRT; 1074 AA.  
AC Q964D1;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Gal/GalNAC lectin Igl2 (Fragment).  
GN IGL2.  
OS Entamoeba histolytica.  
OC Eukaryota; Entamoebidae; Entamoeba.  
OX NCBI\_TaxID=5759;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HM1:IMSS;  
RX MEDLINE=21391855; PubMed=11500468;  
RA Cheng X.J., Hughes M.A., Huston C.D., Loftus B., Gilchrist C.A.,  
RA Lockhart L.A., Ghosh S., Miller-Sims V., Mann B.J., Petri W.A. Jr.,  
RA Tachibana H.;  
RT "Intermediate Subunit of the Gal/GalNAC Lectin of Entamoeba  
RT histolytica Is a Member of a Gene Family Containing Multiple CXXC  
RT Sequence Motifs";  
RL Infect. Immun. 69:5892-5898(2001).  
DR EMBL; AF337951; AAK92362.1; -.  
DR InterPro; IPR000561; EGF-like.  
DR PROSITE; PS01186; EGF\_2; UNKNOWN\_1.  
KW Lectin.  
FT NON\_TER 1  
FT NON\_TER 1074 1074  
SQ SEQUENCE 1074 AA; 116782 MW; 1552E2D714EB450F CRC64;

Query Match 4.2%; Score 222.5; DB 5; Length 1074;  
Best Local Similarity 19.5%; Pred. No. 7.8e-10;  
Matches 194; Conservative 113; Mismatches 361; Indels 325; Gaps 52;  
QY 107 CASGEYLEMKNQVCSKCGEGTYSLSGSIKGFDEWDELPAGFSNIAFPMDTVVGPDSRPGD 166  
Db 13 CAS-----VSNAGACASCDEG-YELKT-----ESGSGTQKCTLKEETCKSAFSYYDG 58  
QY 167 CNNSSWIPRGNVIESNRDDCTVSLIYAVHLKKSQYVFFEQYQVVDNIIFFEFFI--QNDQC 224  
Db 59 SDSNS--PKVCYCENGKESDTSNNNEKCKCKNG-----VDT---CESCLSKDNDKC 105  
QY 225 QE---MDTTDKWKVKLTDNGEWGSHS-----VMLKSGTNILYWRITGI--LMGSKAVKP 273  
Db 106 GECVIGMSTYTTGGQKLCDTVTTDEHAENCVLGTAKDSSSKQCDKCFGMYSLSQGQCTK- 164  
QY 274 VLVKNITIEG--VAYTSECFCPKPGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSD 331  
Db 165 ---KNEKIEKCILOVESSCNQCADGYVINT--EKCKTKYPDHC-SKMNSDKNCGMEGYY 218  
QY 332 FSGSSECTERPPCT---TKDYFQIHPTPCDEEGKTOIMYKWIETKICREDLT--DAI---- 382

Db 219 LNG-----TECKVCTIDNSKDL-----SEGNQCSIYNAEHCECNKRCTVSDGVCVKN 266  
Qy 383 --RLPPSGEKKDCPPCPNPGFYNNGSSSCHP-----CPPGTF--SDGTKE-CR 424  
Db 267 HCRLFSPTENKCTKCDGDGYFLTGAGKCSPLNDGFKTSKTECQKGYYLEKDGDKKRCS 326  
Qy 425 PCPAGTEPALGFYKWNVLPGNMK-----TSCFNVGNSK 459  
Db 327 LCPDPFTECLTSQ---TPVPGKLNRLSAHLTSTDGPCKLPGLCLCSDDDTICY-----K 377  
Qy 460 CDGMNGWEVAGDH-----IQSGAGGSDNDYLILNL-----HIPGFKPPTSMTGATGS 506  
Db 378 CE--NGLTLNGTHCYNFDVKVKVLGTSGNNHQVCKMRGYDQYEQYLNAFKASDN----- 428  
Qy 507 ELGRITFVFETLCS-ADCVLVYFMDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW 565  
Db 429 -----TY-----YCPLKDLVLPYFVSVTGSDNKITIGCVGKDRDV----- 464  
Qy 566 AFQRTNOQDNRRI-----NDMKYIYSITATNAVGVASS--CRACALGSEQSGSSC 616  
Db 465 ----KNDCECNKDIYIPKSVDKASDCVSIK--TKLPSCERAANENICTQCPVGVSHVDSNGK 518  
Qy 617 VPCPPGHYIEKETNOCKECPD-TYLSIHQVYGKEACIPC-----GPGSKNNQDHSVCS 670  
Db 519 CSCGDHAYFD-QNNKQCECPASCSSCYDSSKSVVCSEYENIQGVSTRDKDNECAKCK 577  
Qy 671 DCFYHEKENQILHYDFSNLSSVGLMNGPSFTSKGTYFHFHFNISLCGHEG----- 722  
Db 578 DTPEYKEGLNA-----EDKKKSCAQLNN-----CKEEGHYKISDGF 614  
Qy 723 -----KMALCTNNITDFTVKEIVAGSDDYTNLVGAFCVQCSTIIPSES 765  
Db 615 ITCLECD SAYIVDSOTKECAQCASN-----AFKDNNK-----CQLCSTKKDK 658  
Qy 766 KGFRAALSSQS-IILADTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSSTATTS 824  
Db 659 YGHCASACATACIICEDTNVLAASGSNAQCTVCKDGFYIESPTDGVY---CSPCPAK 714  
Qy 825 CINGRSTAVK--MRCNPTKSGAGVISVPSKCPA-----GTCDCGCT----- 862  
Db 715 CKTKYSADKKEIECVTCTDQSSVDIKPPTCACLTGTVQLENGTCQSCSDLSKYPGCKTT 774  
Qy 863 -----FYFLWESAEA-----CPLCTEHDFHEIEGACKRGFQETLYVWNEPK 903  
Db 775 DTCNVDSRTGIYATECSDGFSGRSPSNCTTCIESNYYPKEG-----EKN 820  
Qy 904 WCIKGISLPEKKLATCETVDFW-----LKVGA 930  
Db 821 GCAK----CDDKCATCSDKDTCCTDPLKIGS 849

RESULT 6  
Q964D2  
ID Q964D2 PRELIMINARY; PRT; 1101 AA.  
AC Q964D2;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Gal/GalNAC lectin Igl1.  
GN IGL1.  
OS Entamoeba histolytica.  
OC Eukaryota; Entamoebidae; Entamoeba.  
OX NCBI\_TaxID=5759;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HM1:IMSS;  
RX MEDLINE=21391855; PubMed=11500468;  
RA Cheng X.J., Hughes M.A., Huston C.D., Loftus B., Gilchrist C.A.,  
RA Lockhart L.A., Ghosh S., Miller-Sims V., Mann B.J., Petri W.A. Jr.,  
RA Tachibana H.;  
RT "Intermediate Subunit of the Gal/GalNAC Lectin of Entamoeba  
histolytica Is a Member of a Gene Family Containing Multiple CXXC

RT Sequence Motifs.";  
RL Infect. Immun. 69:5892-5898(2001).  
DR EMBL; AF337950; AAK92361.1;  
DR InterPro; IPR000561; EGF-like.  
DR PROSITE; PS01186; EGF\_2; UNKNOWN\_1.  
KW Lectin.  
SQ SEQUENCE 1101 AA; 119512 MW; C8B6F5CBDE656AEC CRC64;  
  
Query Match 3.9%; Score 211.5; DB 5; Length 1101;  
Best Local Similarity 20.0%; Pred. No. 7e-09;  
Matches 202; Conservative 104; Mismatches 351; Indels 355; Gaps 59;  
  
Qy 101 KECTFSCASGEYLEMKNQVCSKCGEGTYSLGSIGIKFDENDELPAAGESNIATFMDTVVGPFS 160  
Db 27 REAVPHCAS-----VSNGACTSCDTG-YELTT-----TGNKKTCTLKEDMCKTA 69  
Qy 161 DSRPDGCNNSSWIPRGNVI-----ESNRDDCTVSLIYAVHLKKSGYVFFEYQVVD 210  
Db 70 FSYVDKTNSTN--PKCTYCVNGKEVNTSSHSNGNDKCVCK----- 106  
Qy 211 NNIFF---EFFIQNDQCE---MDTTTDKWKVKTLDNGEWSHSV---MLKSGTNI--- 256  
Db 107 NNVNICESCLLMKDSKCGECIIGMSTTVD-GSKLCDNATEDHAENCVGLLASSTSSKTC 165  
Qy 257 -----LYWRTTGILMGSKAVKPVLVKNITIEG--VAYTSECFPCKPGTFSNKPGEFNCQV 309  
Db 166 DKCEGMY-----SLQGGKCTQ-----KNDKINKILQVENSNOQADG-YSLSTDKKSCNK 215  
Qy 310 CPRNTYSEKGAKECIRCKDDSOFSGS-SECTERPPCTTKDYFQIHPTCD-EEGKTQIMYK 367  
Db 216 FPEHC-SKINGNQCLTCMEGYVLSKTDSKCT---ICT-----VDNPNNLSEGNECSIYN 265  
Qy 368 WIEPKICREDLT--DAI-----RLPPSGEKKDCPPCPNPGFYNNGSSSCHP----- 410  
Db 266 AEHTSCNKRCTVSDGVCVKNHCRFLFSPTENKCTKCDNGYFELTSGTCSPLNYDGFKTA 325  
Qy 411 ---CPPGTF--SDGTKE-CRPCPAGTEPALGFYKWNVLPGNMK----- 449  
Db 326 NRTECENGYYLEKDGDKKRCSLCPDPFTECLTSK---TPVPGKLNLRSSHLTSTDGPCK 381  
Qy 450 -----TSCFNVGNSKCDGMNGWEVAGDH-----IQSGAGGSDNDYLILNL----- 489  
Db 382 LPGLCLCSDDDTICY---KCE--NGLTLNGTHCYNFDTKSVLGTSGNNHQVCKMRGYD 434  
Qy 490 ---HIPGFKPPTSMTGATGSELGRITFVFETLCSADCVL-----YFMVDINRKSTNV 538  
Db 435 QYEQYLNAFKASDN-----TYY-----CPLKDLVLPYFVSVT----- 466  
Qy 539 VESWGGTKEKQAYTHIIFKNATFTFTWAFQRTNOQGD---NRREI-----NDMKIYS 588  
Db 467 ----KGTSDN-----TITIGCVQLRNVSDCECDKHIIPTSIDKASDCVSIIT 511  
Qy 589 ITATNAVGVASSCRACALGSEQSGSSCVPCPPGHYIEKETNOCKECPD-TYLSIHQVY 647  
Db 512 KLPSCERTANGNICTQCPVGVSHVKGDKGKSCGDAHYFDKD-NVCKKCPASCSSCSYDSSK 570  
Qy 648 GKEACTPC-----GPGSKNNQDIHSVCYSDCFFYHEKENQILHYDFSNLSSVGLMNGPSF 702  
Db 571 SKVVCSEYENIQGVTRNKENECACINDGY---KE-----GPNA 607  
Qy 703 TSKGTYFHFHFNISLCGHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFV-----CQ 757  
Db 608 EDKKKSCAQLNN--NCKKEGK-----YEISDGFV--TCLDCCDSAYIVGSQVGAQTQCS 657  
Qy 758 STIIPSESKGFRAALSSQSIIILADTFIGVTVETTLKNINIKEDMFPVPTSQIPDV---HF 814  
Db 658 PNAFKDENNKCOLCSTKQSQYGHCAACATACITCEDINL-----ILTGEKPCCTVCKDG 711  
Qy 815 FYKSSSTAT-----TSC--INGRSTAVKMRC-NPTKSGAGVISVPS-KCPA----- 855  
Db 712 FYQIENATDGVVCSPCPAKCKTKCKYNTTSKKVECVCTCEQRKLDIKAPECACPTGTVQLE 771  
Qy 856 -GTCDCGCT-----FYFLWESAEA-----CPLCTEHDFHEI 884

```
Db 772 NGTCQSCSLSKYPGKKTDCNVDSRTGFIYATECSDGFSGRSPYSNCTTCTKSNNYYPK 831
QY 885 EGACKRGFQETLYVWNEPKWCIKIGISLPEKKLATCETVDFW-----LKVGA 930
Db 832 EG-----EKNCGAK-----CDDKCATCSDKDTCLTLCADPLKVG 865-

RESULT 7
Q9U048
ID Q9U048 PRELIMINARY; PRT; 704 AA.
AC Q9U048;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Variant-specific surface protein.
GN VSP417-6.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AD-1;
RA Ey P.L., Darby J.M.;
RT "A new member of the vsp417 subfamily of variant-specific surface
RT protein (VSP) genes in Giardia intestinalis.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF065606; AAF02907.1;
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR005127; Giardia_VSP.
DR Pfam; PF03302; VSP; 2.
DR SMART; SM00261; FU; 4.
SQ SEQUENCE 704 AA; 71677 MW; 7E5AE1245AD4FD45 CRC64;

Query Match 3.9%; Score 207.5; DB 5; Length 704;
Best Local Similarity 20.8%; Pred. No. 8.2e-09;
Matches 199; Conservative 88; Mismatches 353; Indels 315; Gaps 54;

QY 38 ALAGCOAAWAGDLPSSSRPLPPCQEKDYHFEY-TECDSSGSRWRVAIPNSAVDCSGLPD 96
Db 15 ARAACQADGSGDGSCKTCGVITGQE-----QYCECNGA-----NYA----- 52

QY 97 PVRGKECTFSCASGEYLEM---KNQ-VCSKCG-----EGTYSLGSGIKFDEWDELPA 145
Db 53 PVNG-QCADVNAEGQSKTLCPAKGQKCTQCGGASFMYKDCYS----- 95

QY 146 FSNITATFMDTVVGPS---DSRPDGCNNSWIIPRGNYIESNRDDCTVSLIYAVHLKSGYV 202
Db 96 -----KDTAPGQSMCTQASEGKCTEAAPGYFLNPLRANTKDSVWSCSDTTGFTDSGKT 148

QY 203 FFEYQYVDNNIFFEYFIQNDQCQEMDTTDDKWVKLTDNGEWGSHSVMLKSGTNILYWRTT 262
Db 149 YRGVOH-----CERCDGAA-----LTDA--GGDAKTRCGQD-KYLATT 185

QY 263 GILMGSKAVKPVLVKNITIEGVAYTSECFPKPGTFSNKPGSFN---CQVCPRNTYSEKG 319
Db 186 GTCG-----EG-----CTPDTEFSKEDSDNGKRCFACGDVT---TG 218

QY 320 AKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIPEKICREDLT 379
Db 219 VASCEKCTPPSPDQAKPACTK---CGGNYY--LKTAAD--GTTTC-----VEQSACSPD-S 266

QY 380 DAIRLPPSGEKKDCPPCNPGFYNNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFYK 439
Db 267 FPVENSQSGNR--CVLCGDA-ANGGVDKCAACTP-----TDQGRIAPTITCIACNGYK 317

QY 440 WNVNLPGNMKTSCFNVGNKCDGMNGWEVAGDHQSGAGGSDNDYLILNLHIPGFKPPTS 499
Db 318 -----PSADKTTCEAVSNCKTPGCKACSNEGKENEVCTDCDSSTYL-----TPTS 362

QY 500 MTGATGSELGRITVFETILCSADCVLYFMVDINRKNSTNVVESWGGTKEKQAYTHIFKNA 559
Db 363 Q-----CIDSCA-----KIGNYYGATEGAKK-----ICKEC 388
```

```
QY 560 TFTFTWAFQRTNQDNRFRINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPC 619
Db 389 TA-----ANCKTCDGQ-----GQCACSDGFYKNGDACSPC 419

QY 620 PPGHYIEK-----ETNQCKECPDFTYLSIHQVYKAEACI-PCGPGSKNNQDHSVYSDCF 673
Db 420 ---HESCKTCSAGTASDCTECPGKALR---YGGDGTGKTCGEGCTTGQSGAC----- 467

QY 674 FYHEKENOILHYDFSNLSSVGLMNGPSTFTSKGTYKFFHFFNLSLCGHEGKKMALCTNNIT 733
Db 468 -----KTCGLTIDGASYCSECATTTEYPQNGVC---APKASRATPTCN 507

QY 734 DFTVKEIVAG--SDDYTNLVGAFVCQSTIIIPSESKGFRAALSSQSIIILADTFIGVTVETT 791
Db 508 DSPIQNGVCGTCADNYFKMNGG--CYETV---KYPGKTVICISAPN-----GGTCQKA 554

QY 792 LKNINIKEDMPVPVPTSQIPDVHFFYKSSSTATTSCTING--RSTAVKMRCN---PTKSGAGV 846
Db 555 ADGYKLDSGTLTVCSEGGKEC-----TSSTDCTTCLDGYVKSASACTKCDSSCETCNGA-- 608

QY 847 ISVPSKCPAGTCDGC-TFYFLNESAE-ACPLCTEHDFHEIEGACKRGFQETLYVWNEPKW 904
Db 609 -----ATTCACATGYKKTASGEGACTSC-ESDSNGVTGI-----KG 644

QY 905 CIKGISLPEKKLAT--CETVDFWLKVGAGVGAFATVL-----LVALTCYFW 948
Db 645 CL-NCAPPSSTGSLVLCYLIQNTKNSGLSAGAIAGISVAVIVVAGLVGLCWVF 698

RESULT 8
Q9XTJ7
ID Q9XTJ7 PRELIMINARY; PRT; 709 AA.
AC Q9XTJ7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Variant-specific surface protein (Fragment).
GN VSP417-6.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRIS-136;
RX MEDLINE=99026095; PubMed=9806870;
RA Ey P.L., Darby J.M.;
RT "Giardia intestinalis: conservation of the variant-specific surface
RT protein VSP417-1 (TSA417) and identification of a divergent homologue
RT encoded at a duplicated locus in genetic group II isolates.";
RL Exp. Parasitol. 90:250-261(1998).
DR EMBL; U89266; AAD03483.1;
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR005127; Giardia_VSP.
DR InterPro; IPR001239; Kazal_inhib.
DR Pfam; PF03302; VSP; 2.
DR PRINTS; PR00290; KAZALINHBTR.
DR SMART; SM00261; FU; 3.
FT NON_TER 709
SQ SEQUENCE 709 AA; 71516 MW; 3512BB844B38D134 CRC64;
```

Query Match 3.8%; Score 203; DB 5; Length 709;  
Best Local Similarity 20.8%; Pred. No. 2e-08;  
Matches 162; Conservative 70; Mismatches 272; Indels 273; Gaps 40;

```
QY 38 ALAGCOAAWAGDLPSSSRPLPPCQEKDYHFEYTECDSSGSRWRVAIPNS--AVDCSGLP 95
Db 15 ARAACQADGSGSAGS-----CKTCG---VAIGQEYQYCECNGAN 50

QY 96 -DPVRGKECTFSCASGEYLEM---KNQ-VCSKCGEGTYSLGSGIKFDEWDELPAFNSIA 150
Db 51 YAPVNG-QCADVNAEGPSKTLCPAKDQKCTQCGNSFMLAGGC-YSSGEGLPGHSLCLS 108
```



QY 151 TFMDTVVGPSDSRPGCGNNSSWIPRGNIESNRDCTVSLIYAVHLKKSQYVFEEYQY-- 208  
Db 109 SGGDGV-----CTEAAPGYFLNPLRANTKDSVVVSCSDTAGFTDSGKTYRGVQYCE 158  
QY 209 -VDNNIFFE-----FFIQNDQCQEMDTTIDKWKVL-TDNGE--WGSHSV 249  
Db 159 RCDGAVLTDAAAGDAKCTRCGENKYLATTGTCGEGCTPDTEFSKEDSDNGKRCFACGDV- 217  
QY 250 LKSTNILYWRITGILMGSKA-----VKPVLV-----NITIEGVAYTSECFPCPKP 295  
Db 218 -----TTGVASCEKCTPPSPDQAKPACTKCGGNNYLKTAADGTTTCAEQSACSP 266  
QY 296 GTF--SNKPGSFNCQV-----CPRNTYSEK-----AKECIRCKDDSQFSG---- 334  
Db 267 DSFPVENSQSGNRCVLCGDAANGVDKCAACTPADKGRAAPAVTCTACTDGYKPSADKTT 326  
QY 335 -----SSECTERPPCTTKDYFQIHTP-----CDEEGKTQIMYKWKIE--PKI 373  
Db 327 CEAVSSCKTPGCKACSNEGKENEVCTDCDGSYTLTPTSQIDSCAKVGNYYGAIEGAKKL 386  
QY 374 CREDLTDAIRLPPSGEKKDCPPCNPFGYNNGSS-----SCHPCPPGTFSDGTKECRPCP 427  
Db 387 CKE-CTAANCKTCDGQGR-CQTCSDGFYKNGDACSPCHESCKTCSAGTASD-----CTECP 440  
QY 428 AGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAG----- 470  
Db 441 TGKALRYGDD-----GTKGTC-GAGCATGGQSGACKTCGLTIDGASYCSECATATE 490  
QY 471 -----DHIQSGAGGS-DNDYLILN-----LHIPG----FKPP-- 497  
Db 491 YPQNGVCAPKASRATPTCNDSPIONGCGTCANSYFKMNGGCGYETVKYPGKTVCISAPNG 550  
QY 498 -TSMTGATGSELGRITFVFETLCSADCVLVFMVDINRKSTNVVSW-----GGTKEKQAYT 552  
Db 551 GTCQKAADGYKLDSTL--TVCSECG-----KECTSSSTDCTTCLDGYVKSASACT 598  
QY 553 HIIFKNATFTTWFQFRTNQDNRFRINDMVKIYSITATNAVVDGVASSCRACALGSEQS 612  
Db 599 KDFSCET-----CNGAATTCACATGYKYKT 624  
QY 613 GSSCVPCPGPHYIEKETNQCKECPDPTLTSIHQVYGKEACIPCGPGSKNNQDHSVY 669  
Db 625 ASG-----EGACTSCESDS-----NGVTGIKGLCNCAPPS-SNTGSVLGY 663

RESULT 9  
O97444 PRELIMINARY; PRT: 709 AA.  
AC O97444;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Variant-specific surface protein 417-4 (Variant-specific surface protein type 4 TSP11/TSA417-like).  
GN VSP417-4.  
OS Giardia lamblia (Giardia intestinalis).  
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.  
OX NCBI\_TaxID=5741;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AD-1;  
RA Ey P.L., Darby J.M., Mayrhofer G.;  
RT "A new locus (vsp417-4/A-I) belonging to a subfamily of tsa417-like variant-specific surface protein (VSP) genes in Giardia intestinalis.";  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 476-705 FROM N.A.  
RC STRAIN=VARIOUS STRAINS;  
RA Ey P.L., Darby J.M.;  
RT "Identification of a subset of tsa417-like genes within the variant-specific surface protein (VSP) gene family of Giardia intestinalis.";  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF065168; AAD04339.1; -.  
DR EMBL; U89153; AAD05041.1; -.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR005127; Giardia\_VSP.  
DR Pfam; PF03302; VSP; 2.  
DR SMART; SM00181; EGF; 1.  
DR SMART; SM00001; EGF\_Like; 3.  
DR SMART; SM00261; FU; 5.  
SQ SEQUENCE 709 AA; 72664 MW; 6E71F27D2F367F52 CRC64;  
  
Query Match 3.8%; Score 201; DB 5; Length 709;  
Best Local Similarity 19.6%; Pred. No. 3e-08;  
Matches 172; Conservative 79; Mismatches 297; Indels 328; Gaps 45;  
  
QY 91 CSGLPDPVRGKECTFSCASGEYLEMKNQVCSCGEGTYSLSGSIKFD---EWDELPAGFS 147  
Db 19 CTQEADDGKCKTCGVTTIGQEY-----CSECNGANYAPVNGACEDVETQADKKALCKA 71  
QY 148 NIATFMDTVVGPDSRDPDGCNNSSWIPRG-----NYIESNRDD 185  
Db 72 HASGACTTCGGNSFMYKDCGYSSGEGLPGHSLCLSSDGDGVCTEAAPGYFLNPLRANTKD 131  
QY 186 CTVSLIYAVHLKKSQYVFEEYQYVDNNIFFEFTIQNDQCQEMDTT----- 231  
Db 132 SVVSCSDTTGFTDSGKTYRGVY-----CERCDGAALTDAAAGDAKCTRC 176  
QY 232 --DKWKV--LTDNGEWGSHSVMLKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAY 286  
Db 177 GQDKYLKDNCTVDKAQCDSG-----TN-----KFVAVDISEN 209  
QY 287 TSECFPCPKPGTFSNKPGEFNCQVCPRNTYSEKAK-ECIRCKDDSQFSGSE---CTERP 342  
Db 210 GNKCVSCSDNL--NGGVANCDTC--SYDEQSKKIKCTKTDNNYLKTTSEGTSCVQKD 263  
QY 343 PCTTKDYFOIHTPCDEEGKTIQIMYKWIPIKICREDLTDALRPPSGEKKDCPPCNPGEFYN 402  
Db 264 QC--KDGf--FPKDDSS-----AGNK--CLPCNDS--T 288  
QY 403 NGSSSCHPCP--PGTFSOGTKECRCPCPAGTEPALGFYKWNVLPGNMKTSFNVGNSKC 460  
Db 289 DGIANCATCALVSGRGAALVTCSTACTDGYKPSAD-----KTTCEAVSNCKT 335  
QY 461 DGMNGWEVAGDHIQSGAGGSNDYILNLHPIPGFKPPTSMTCATGSELGRITFVFETLCS 520  
Db 336 PGCKACSNEGKENEVCTDCDSSTYL-----TPTSQ-----CI 367  
QY 521 ADCVLVFMVDINRKSTNVVSWGGPKKEKQAYTHIIFKNATFTTWFQFRTNQDNRRTI 580  
Db 368 DSCA-----KIGNYYGATEGAKK-----ICKECTA----ANCKTCDGQ----- 401  
QY 581 NDMVKIYSITATNAVVDGVASSCRACALGSEQSGSCVCPGPHYIEK-----ETNQCKEC 635  
Db 402 -----GQCQACSDGFYKNGDACSPC---HESCKTCSAGTASDCTEC 439  
QY 636 PPDTYLSIHQVYGKEACI-PCGPGSKNNQDHSVYSDCFYHEKENQILHYDFSNLSSVG 694  
Db 440 PTGKALR-----YGDDGKTGTCGAGCATGTGAGAC-----KTCG 473  
QY 695 SLMNGPSFTSK---GTKYFHFENISLCGHEGKKMALCTNNITDFIVKEIV---AGSDDYT 748  
Db 474 LTIDGASYCSECATATEY-----PQNGVCSST---TVRAAATCKAGS----- 512  
QY 749 NLVGAFVQCSTIIPSESKGFRAALSSOSIILADTFIGVTVETT-LKNINIKEDMFPV-PT 806  
Db 513 --VAKGMCNSC-----TNGFL-----RMNGGCYETTKFKPKNVCEEAAPAGDT 553  
QY 807 SQIPDVHFFYKSSSTATTSINGRSTAVKM-RCNPTKSGAGVIS-----VP-----SKCP 854  
Db 554 CQTP-ADGYKLNNGALITCSAGCKTCTSQDQCDTCKAGYAKTGNGTKKCVPCATGCSECN 612  
QY 855 AGTCDGCTF---YFLWESAECPLCTEHPHEIEG 886

```
Db 613 ADDATKCTVCAAGYYL--SKEKCIADKSDGGSITG 646

RESULT 10
Q95WU1 PRELIMINARY; PRT; 667 AA.
AC Q95WU1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Variant-specific surface protein WB/9B10-B.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RA Lujan H.D.;
RT "Identification of a novel variant-specific surface protein of Giardia intestinalis.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF293416; AAK97086.1; -
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR005127; Giardia_VSP.
DR Pfam; PF03302; VSP; 2.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
SQ SEQUENCE 667 AA; 68451 MW; 769372C32C689546 CRC64;

Query Match 3.7%; Score 199.5; DB 5; Length 667;
Best Local Similarity 19.6%; Pred. No. 3.7e-08;
Matches 149; Conservative 70; Mismatches 275; Indels 265; Gaps 39;

QY 37 WALAG--CQAAWAGDLPSSSSRPLPPCQEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGL 94
Db 5 FVLAGVLVQIAWAGKATERAAQ----CADN-----TNCAEEACNVLIGGKLYCSRCNTG 54

QY 95 PDPVRG-----KE-CTFSCASGEYLEMKNQVCSKCGEGTYSLGSGIKFDEWDELPAFESNI 149
Db 55 FVPINGQCADKEGATDQCKDGGSGDADQTCCGCAEQTFMYKGGCY--EAAQQPG----- 107

QY 150 ATFMDTVVGPDSRDPDGCNNSWIPRGNVIESNRDDCTVSLIYAVHLKKSQYVFEYQYV 209
Db 108 ----QTMCOADA--GVCTQAA---QGYFVPPGADASHQSVIPCG--DEEGIT-----V 150

QY 210 DNNIFFEFFFIONDQCEMDTTTDKWKVLTNDNGEWSHVMKSGTNIILYWRRTTGILMGSK 269
Db 151 KNDKKYKGVLHCTRCYAPTEAADANAKAATCTACGDSKIV-----K 191

QY 270 AVKPVLVKNITIEGVAYTSECFPCPKPGTF-----SNKPGSFNCQVCPRNTYSEKGAK 322
Db 192 TAKDSATSCVTEEECTGKTCKTCAEGTSDGCATCEKGADGAVACKTCGSNKKVQPNKKG 251

QY 323 CI-RCKD--DSQFSGSSECTE-----RPPCTTKDYFQIHPT-----CDEEGKTQIMYKW 368
Db 252 CIAKCPETVSAEKDGVCEVGEYVPDNAGTGTCKKPDQCNTPPGCKTCS----- 300

QY 369 IEPKICREDLT---DAIRLPPSGE-----KK----- 391
Db 301 -EPKTSKEVCTECEDPKALTPTGQCIYGEHLEGYEGTSEGGKACKKCEVENCLLCNG 359

QY 392 --DCPPCNPGFYNNG-----SSSCHPCPPGTFSDGTKECRP-----CPAGT 430
Db 360 QQCETCKDGYKSGAACAKCNTSCKTCANGN--SNGCTSCPEKQVLSYEGEGTGTCKPGC 418

QY 431 EPALGFEYKWNVLPGNMKTSFCNV--GNSKCDGMNGWEVAGDHQSG----- 476
Db 419 KPVSGGK-----DGTCKSCDLNIDGTSYCSACN---VGTPEPENGVCVKKSARTASCO 468

QY 477 -----AGGSDNDYLILN-----LHIPG-----FKPPTSMTGATGSELGRITFVF 515
Db 469 AEPSNGVCGTCARGFFRMNGGCVETTKLPKGSVCEEVASAGDTCQTPADGYKLNNGALI- 527

QY 516 ETLCSDCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTWAFQRTNOGOD 575
||| |
```

```
Db 528 --TCSAGC-----KTCTSQD 540

QY 576 NRRFINDMVKI-YSITATNAVDGVASSCRACALGSEQ----SGSSCVPCPPGHHYIEKETN 630
Db 541 Q-----CDTCKAGYAKTGGN-----TKKCVPCATGCSSECNADDAKCTVCAAGYYLSKE-- 589

QY 631 QCKECPPTDYLSTHQQVYVGKEACIPCGPGSGSKNNQDHSVCY 669
Db 590 KCIADCKSDGGSF---TGVANCANCAPPT--NNKGPVLGY 624

RESULT 11
Q8WRD2 PRELIMINARY; PRT; 3087 AA.
AC Q8WRD2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cysteine repeat modular protein 1 PbCRM1.
OS Plasmodium berghei (strain Anka).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ANKA;
RA Thompson J.;
RT "Plasmodium berghei Cysteine Repeat Modular Protein 1: PbCRM1.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF449196; AAL47156.1; -
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001664; IF.
DR InterPro; IPR000001; Kringle.
DR InterPro; IPR000589; Ribosomal_S15.
DR Pfam; PF00051; kringle; 1.
DR SMART; SM00130; KR; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_3.
DR PROSITE; PS00226; IF; UNKNOWN_1.
DR PROSITE; PS50070; KRINGLE_2; 1.
DR PROSITE; PS00362; RIBOSOMAL_S15; UNKNOWN_1.
SQ SEQUENCE 3087 AA; 362575 MW; 38BDD25AB65FD43D CRC64;

Query Match 3.7%; Score 197.5; DB 5; Length 3087;
Best Local Similarity 20.0%; Pred. No. 4.7e-07;
Matches 197; Conservative 127; Mismatches 322; Indels 337; Gaps 55;

QY 102 ECTFSCASGEYLEMKNQVCSKCGEGTYSLGSGIKFDEWDELPAFESNIATFMDTVVGPSP 161
Db 1317 EFSIACEDGHY--HKHNKCHPCEGYNNLNEIK-----QNN 1351

QY 162 SRPDGC---NNSSWIPRGNVIESNRDDCTVSLIYAVHLKK-----SGYVFEYQY 208
Db 1352 EKYKKCTACGQNRRTTLDKKEFEKN---CLCDLGYE-YIKNPNPNRNFICSPCVGEYKD 1407

QY 209 VDNNIFEEFFIONDQCEMDTTTDKWKVLTNDNGEWSHVMKSGTNIILYWRRTTGILMGS 268
Db 1408 V-----ISNELCK-----GIVCKKNASYL-----ILDK 1430

QY 269 KAVKPVLVKNITIEGVAYTSECFPCPKPGTF--SNKPGSFNCQVCPRNTYSEKGAKCIRC 326
Db 1431 KSVDP-----SQCL-CNGGYLYRYDKNNNEICTKINNHY-----C 1465

QY 327 KDSQFSGSSEC-----TERPPCTTKDYFQIHPTCDEEGKTQIMYKWIPEKICRE----- 376
Db 1466 PDNDNY--YKKCPIHNKTEEIKRT--DFETINSLCEEGYEPINMNKIKDTSRDYHYNN 1521

QY 377 -----DLTDAIRLPPSGEKKDCPPCNPGFYNNGSSS--CHPCPPG--TFSDGTK---E 422
Db 1522 IFVTKYKNLSYIV-----NDNNICMECNLGFYKNNISSEKCIKPRSSSTTTTYSKSIQN 1576

QY 423 CRPCPAGTEPALGFEYKWNVLPGNMKTSCTSC-----FNVGNS-KCDGMNGWEVAGDHI- 473
||| |
```

Db 1577 CNSCHKG-----YYK-----DQKQVCSKCLPNHFCVGSQKNDKNNISQYAGDAII 1622  
QY 474 -----QSGAGGSDNDYLILNLH-----IPG 493  
Db 1623 CPNYSVTLQPYENNISFKNCLCIKGYEKNFQDFYNNHCKKAPLNFYKDTISNDLSIP- 1681  
QY 494 FKPPSTMTGATGSELGRITPVFETLCSADCVLYFMVDINRKSTNVVESWGGTKKEQAYTH 553  
Db 1682 -CPNSTITLNTGA-----TSIYNICIDKG---FFYDTVSYSCVNCPHGY-YCSEKDMTTK 1731  
QY 554 I-----IFKNATFTTFAFORTNQGDNRFRINDMWKISITATNAVGVASSCRACALG 608  
Db 1732 LAPPIKCPKNYTTIYKGSYNISHCVCESGYTVNTVIVEHS-SENGIIESHESAKNKKIKLQ 1790  
QY 609 SEQSSSCVPCPPGHIYEKTN-QCKECPDT-----YLSIHQVYVGEACI 653  
Db 1791 AKNRTSVCKCPOSSYKSKISNEQCHKCPKNSKTLKDFNNSDIFFCLCTMGYYTDKKECK 1850  
QY 654 PCG-----PGSKNNQDHSVYCSDCFYHEKENQILHYDFSNLSSVGLMNGPSTSKGT 707  
Db 1851 PCWFNKLYCEGEKIYQIETIYDEIINLIQYIRLLSDISTKKSFINIIN--EYINKGL 1908  
QY 708 KYHFHFNISLCGHEGKKMALCTNITDFTVKEIVAGSDDYT---NLVGAfv-----CQS 758  
Db 1909 -ISEITHISMMEKSRKKSg---NNIQ--SIRKAKKDSEIYAVNIYNRLGNIITKDRELKT 1962  
QY 759 TIIPSESGFRAALSSQSIIADTFIGTVETTL---KNINIKEDMFPVPTSQIPDVHFF 815  
Db 1963 PKINNIQKIELAKSIENKIIAN-----INSKLTLLSEGKNIDTD-----TSIL--ANLF 2011  
QY 816 YKSS-----TATTSCIN-GRSTAVKMRNPTKSGAGVISVPS----KCPAGTCDGCTFYF 865  
Db 2012 YKSSENIYIKHQKLINCQRTNVIPL-----GVDSSQNFDCKCKKG-----YY 2055  
QY 866 L-----WESAECAPLCTEHDFHEIEG-----ACKRGFQETL 896  
Db 2056 LEDRIVLKNICKPCPEGTFKNDYDGVKKCISCPPKSTSIKSIYPNHCCKNGF---F 2112  
QY 897 YVWNEPKWCIKGISLPEKKLATC 919  
Db 2113 YSKDTCLELEG-----ATC 2127

RESULT 12  
Q99ND0  
ID Q99ND0 PRELIMINARY; PRT; 5374 AA.  
AC Q99ND0;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE ZAN (Zonadhesin).  
GN ZAN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SV;  
RX MEDLINE=21138439; PubMed=11239002;  
RA Wilson M.D., Riener C., Martindale D.W., Schnupf P., Boright A.P.,  
RA Cheung T.L., Hardy D.M., Schwartz S., Scherer S.W., Tsui L.-C.,  
RA Miller W., Koop B.F.;  
RT "Comparative analysis of the gene-dense ACHE/TFR2 region on human  
RT chromosome 7q22 with the orthologous region on mouse chromosome 5.";  
RL Nucleic Acids Res. 29:1352-1365(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SV;  
RA Cheung T.L., Wilson M.D., Koop B.F., Hardy D.M.;  
RT "Genomic Basis of Inter- and Intra-species Variation in Zonadhesin  
RT Domain Structure.";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

CC -I- SIMILARITY: CONTAINS 3 MAM DOMAINS.  
DR EMBL; AF312033; AAK28824.1; -.  
DR EMBL; AY046056; AAL04416.1; -.  
DR MGD; MGI:106656; Zan.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR003645; FOLN.  
DR InterPro; IPR000998; MAM\_domain.  
DR InterPro; IPR003328; Tila\_Cysrich.  
DR InterPro; IPR002919; TIL\_Cysrich.  
DR InterPro; IPR001007; VWF\_C.  
DR InterPro; IPR001846; VWF\_D.  
DR Pfam; PF00629; MAM; 3.  
DR Pfam; PF01826; TIL; 25.  
DR Pfam; PF02345; Tila; 25.  
DR Pfam; PF00094; vwd; 4.  
DR SMART; SM00001; EGF\_like; 1.  
DR SMART; SM00274; FOLN; 21.  
DR SMART; SM00137; MAM; 3.  
DR SMART; SM00214; VWC; 25.  
DR SMART; SM00216; VWD; 4.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
DR PROSITE; PS01186; EGF\_2; 18.  
DR PROSITE; PS50060; MAM\_2; 3.  
KW EGF-like domain; Glycoprotein.  
SQ SEQUENCE 5374 AA; 579536 MW; 90D2D8CFE5DE24EB CRC64;

Query Match 3.6%; Score 193.5; DB 11; Length 5374;  
Best Local Similarity 19.3%; Pred. No. 2.3e-06;  
Matches 233; Conservative 103; Mismatches 387; Indels 487; Gaps 64;

QY 23 RGRSPWPSPAWI-----CCWALAGCQ-----AAWAGDLPSSSSR-----PLPPCOE 63  
Db 3122 QGALIPGKWTWITSGCTQSCNCTGGAIQCONFCQPLKTYCKDLKDGSSNCTNIPL-----Q 3177  
QY 64 KDYHFEYTECDSSGSRWRVAIPNSAVDCSGL--PDPVRGKECTFSC-ASGEYLEMKNQ- 118  
Db 3178 CPAHSRYTNCLPS-----CPPSCLDPEGLCEGTSPKVPSTCREGICQPGYLMHKNK 3230  
QY 119 -----VCS-----KCGEGTYSLGSGIKFDE 138  
Db 3231 VLRIFCGCKNTQGAFIGADKTWISRGCTQSCTSAGAIHCRNFKCPSGTCKNGDNGSSN 3290  
QY 139 WDELPAGFSNIATFMDTVVGPDSRPGDCNNSSWIPRGNVIESNRDDCTVSLIYAVHLKK 198  
Db 3291 CTEITLQCPTNSQFTDCL---PSCVPSCSNRCEVTSVPSSCREGC-----LCN 3337  
QY 199 SGYVFFEQYVDNNIFFEFFFIONDOCEMDTTTDKWVKLTDNGEWGSHSVMLKSGTNILY 258  
Db 3338 HGFVFSE-----DKCVPRTQCCKDARGAIIIPAG----- 3366  
QY 259 WRTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCPKPGTFSNKPFSFNCO--VCPRNTYS 316  
Db 3367 -----KTWTSKGTQSCACV-----EGNIQCNFQCPPEY- 3397  
QY 317 EKGAKECIRCKDDSDQFSGSSECTE---RPPCTTKDYFQIHTPC-----DEEGKTOI 364  
Db 3398 -----CKDNSE--GSSTCTKITLQCPAHTQ-----YTSLPSCLPSCLDPEG----- 3437  
QY 365 MYKWIEPKI---CRE-----DLTDAI-RLPPSGEKKDCPPCNP 399  
Db 3438 LCKDISPKVPSTCKEGCVQSGYVLSNDSKCVLRAECDKQAQGALIPAGKTWTSPGCTQS 3497  
QY 400 -FYNNGSSSCH--PCPPGTF---SDGPKECR---PCPA----- 428  
Db 3498 CACMGGAVCQSSQCPPTGYCKDNEGNSNCAKITLQCPAHSLEFNTCLPSCLPSCLDPDG 3557  
QY 429 ---GTEPALCFEYKWNWVLPGNMKTSC-----FNVGNSKCDGMN----- 464  
Db 3558 LCKGASPK-----VPSTCKEGCICQSGYVLSNNKCLLRNRCCKDAHGALIPEDK 3607  
QY 465 -----GWEV-----AGDHIQSGAGGSDNDYLI-----LNLHI----- 491  
Db 3608 TWVSRGCTQSCVCTGGSIQCLSFQCPPPGAYCKDNEDGSSNCARIPPOCPANSHYTDCEFP 3667



Qy 492 --PGFKPPTSMGTATGSELGRITFVFETLCSADCVLY--FMVDINRKSTNVVSWGGTKE 547  
Db 3668 CPPSCSDPEGHCEASGRV-----PSTCREGLCNPGFVLD-RDKCVRVEC--GCKD 3717  
Qy 548 KOAYTHIIFKNATFTFTWAFQ-----RTNQ-----GQDNRFFINDMVKI--YS 588  
Db 3718 AQG--ALIPSGKTWTSPGRTQSCACMGVVCQSSQCQPPGTYCKDNEDGNSCAKITLQC 3775  
Qy 589 ITATNAVDGVASSCRAC---AL--GSEQSGSS-----CVPCPPGHYIEKE----- 628  
Db 3776 PTHSNYTDCLPFLPCLPSCLDPSALCGGTSKGPSTCKEGCV-CQPGYVLDKDKCILKIECG 3834  
Qy 629 -----TNQCK--ECPPDTVLSIHQVYGK----- 649  
Db 3835 CKDTQGAVIPAGKTLSTGCIQSCACVEGTIQCNFQCQPPGTYCNHNNNCAKIPLQCPAH 3894  
Qy 650 ---EACIPCGPGSKNNQDHS-----VCYSDCFYHEKENQILHYDFSNL 690  
Db 3895 SHFTSCLPSCPSPSCANLDGSCQTSKVPSTCKEGCLCQPGYFLNNGKCVLQTHCDCKDA 3954  
Qy 691 SSVGSLMNGPSFTSK-----GTYFYHFFNISL--CGHEGKKMALCTNNITDFTV 737  
Db 3955 EG-GLVPAGKTWTSKDCTQSCACTGGAQCQNFQCPLGTYCKDSGDGSSNCT-KIHKGAM 4012  
Qy 738 KE--IVAGS-----DDYTNLVGAFVCQSTIIPSESKGFRAALSSQ 775  
Db 4013 GDGVLMAIGIRALQCPAHSHFTSCLPSCPSPSCNLDGSCVESNFAPSVCK--KGICICQP 4070  
Qy 776 SIILADTFIGTVVETTLKNINIKEDMEFPVPTSIQPDVHFFYKSTATTSCINGRSTAVKM 835  
Db 4071 GYLLNNDKCVLRIOCGCK--DTQGLIPAGRTWIS-----SDTKSCSCMGGTIOCRDF 4122  
Qy 836 RCNP-----TKSGAGVISVPSKPCAGTCDGCTFYFLWESAACP-LCTEHDHF----- 882  
Db 4123 QCPPGTCKESNDSSRTCAKIPLQCPAHS-----HYTNCLPACSRSCDTLDGHCEGTS 4175  
Qy 883 -EIEGACKRG 891  
Db 4176 PKVPSCKREG 4185

RESULT 13  
Q9U021  
ID Q9U021 PRELIMINARY; PRT; 719 AA.  
AC Q9U021;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Variant-specific surface protein.  
GN VSP417-7.  
OS Giardia lamblia (Giardia intestinalis).  
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.  
OX NCBI\_TaxID=5741;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AD-1;  
RA Ey P.L., Darby J.M.;  
RT "Comparative analysis of the VSP417 subfamily of variant-specific  
RT proteins in Giardia intestinalis.";  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF189719; AAF04387.1; -.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR005127; Giardia\_VSP.  
DR Pfam; PF03302; VSP; 2.  
DR SMART; SM00181; EGF; 2.  
DR SMART; SM00001; EGF-like; 1.  
DR SMART; SM00261; FU; 4.  
SQ SEQUENCE 719 AA; 7388 MW; 83BE706BACE7F977 CRC64;

Query Match 3.6%; Score 192.5; DB 5; Length 719;  
Best Local Similarity 20.9%; Pred. No. 1.6e-07;

Matches 144; Conservative 56; Mismatches 206; Indels 283; Gaps 39;  
Qy 90 DCSGLPDPVRGKE---CTFSCASGEYLEMKNQVC-----SKCGEGYSLGSGIKFDEWDE 141  
Db 159 ECSG-SELTSQDGTAKCT-KCGASKYL--KDNACVDNAEACGKGYFGK-----PD 205  
Qy 142 LPAGESNIATFMDTVVGPDSRSPDGCNNSSWIPRGNYESN--RDDCTVSLIYAVHLKKS 199  
Db 206 AAAGNKCIA-----CTDQSDGGTAGCAECMY-----DSNTRKAICTKCTDYLRKKAD 253  
Qy 200 GYVFFEYQYVDNNIFFEFFFIONDQC---QEMDYYTDKWKVLTNDGEWSHVMKSGTNI 256  
Db 254 G-----TTECVAANECDDTQKGFYKVVVDSTN-GNKCVCSCADGAGL 292  
Qy 257 LY-----WRTGILMGSKAVKPVLVKNITIEGVAYTSECFCPKPG-----TFSNKPGSF-- 305  
Db 293 AVGTDGAWK--GVDGCAKCIKPADINTPT-----KDECKPGYEISTDKTKCTSTAP 342  
Qy 306 -----NCQVCPRNTYSEKGAKEC-----IRCKDDSQFSGS----- 335  
Db 343 PDCPIENCKVCSED---KRACEECNSNNVLTPTRMCIDCKKIGNYYTSSNANNKLICKE 399  
Qy 336 -----SECTERPPCTTKDYFQIHPTPCDEEGKTQIMYKWIET-KICREDLTDAILPPSG 388  
Db 400 CAVANCKECENTGTCKT-----CDG-----FYKSSECKACDSN---CKTCNGG 441  
Qy 389 EKKDCPPCNPGEF---YNN-----GSSSCHPCPPPGTFSDGTEKCRPCPAGT 430  
Db 442 TSADCTKCLSGAVLKYGNDGKTGTCGAGCATGTGAGACKTC--GLIIDGTSYCSECAVET 499  
Qy 431 EPALG-----FEYKWNVNLPGNNMKTSCFN-----VGNKSCDGMN 464  
Db 500 EYPQGGVCSSTTVRAAATCKAGSVAKG-MCNSCTNGFLRMNGGCYETTKFPGKSVCEEA 558  
Qy 465 GWEVAGDHIQSGAGG---SDNDYLILNLHIPGPKPPTSMTGATGSELGRI-TFVFETLCS 520  
Db 559 S---AGDTCQKEAPGYHLNNDLVTCs---PGCKTCTSNVTCTACMEGYVKTSDCAKCA 612  
Qy 521 ADCVLYFMVDINRKSTNVVSWGGTKEKQAYTHIIFKNATFTTWFQRTNQGDNRRFI 580  
Db 613 AGC-----ATCT----- 619  
Qy 581 NDMVKIYSITATNAVDGVASSCRACALGSEOSGSSCVPCPPGHYIEKFTNQCKECPDPT 640  
Db 620 -----GSTTACDTCSTGYKSGTTCVSC-----TESNSDK----- 649  
Qy 641 LSIHQVYVYGEACIPCGPGSKNNQDHSVCY 669  
Db 650 ----TITGVANCAXCAP-PLNNKGSVLCY 673

RESULT 14  
O45614  
ID O45614 PRELIMINARY; PRT; 2759 AA.  
AC O45614; P91824;  
DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE T22A3.8 protein (Fragment).  
GN T22A3.8.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Barlow K.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA McMurray A.A.;  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL008585; CAA15432.2; -.

```
DR EMBL; 281125; CAA15432.2; JOINED.
DR EMBL; 281125; CAB03385.2; -.
DR EMBL; AL008585; CAB03385.2; JOINED.
DR HSSP; P02468; 1TLE.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR001886; LamNT.
DR Pfam; PF00052; laminin_B; 2.
DR Pfam; PF00053; laminin_EGF; 17.
DR Pfam; PF00054; laminin_G; 1.
DR Pfam; PF00055; laminin_Nterm; 1.
DR PRINTS; PR00011; EGFLAMININ.
DR ProDom; PD002082; LamNT; 1.
DR ProDom; PD003031; Laminin_B; 2.
DR SMART; SM00180; EGF_Lam; 14.
DR SMART; SM00281; LamB; 2.
DR SMART; SM00282; LamG; 2.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_11..
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
FT NON_TER 2759 2759
SQ SEQUENCE 2759 AA; 305515 MW; 081F737000C63541 CRC64;

Query Match 3.6%; Score 192.5; DB 5; Length 2759;
Best Local Similarity 18.8%; Pred. No. 1.1e-06;
Matches 205; Conservative 96; Mismatches 326; Indels 461; Gaps 59;

QY 39 LAGQAAWAGDLPSSSSRPLPPC-----QEKDY-----HFEYTECDSSGS 78
| | : | | | | | | | | : | | : | |
Db 819 LGACEQC--ECPSLDLNPPECISTELAVLGSVASNEDNVINCPLGYEGNKCEY--- 872
| | : | | | | | | | | : | | : | |

QY 79 RWRVAIPNSAVDCSG--LPDPVRCK--ECTF-----SCASGEYLEMKNOVCSCK- 123
| | : | | | | | | | | : | | : | |
Db 873 -----CSDGFFEDPLTGKCIECTCNGNIDPMSGICNDS-----ETGKCLKCI 914
| | : | | | | | | | | : | | : | |

QY 124 GEGYSLGSGIKFDEWDELPAFESNIATFMDTVVGPSDSRPDGCNNSSWIPRGNYIESNR 183
| | : | | | | | | | | : | | : | |
Db 915 GHTTGDSCESCKEHHW-----GNAQLHT-----CKPCGCHTQGA VNPQCSENGE 959
| | : | | | | | | | | : | | : | |

QY 184 DDCVTSLIYAVHLKSGYVFFYQYVDNNIFFEFFFIONDQCEMDTTTDKWKVLTDNGEW 243
| | : | | | | | | | | : | | : | |
Db 960 CEKENYIGA-----QCDRCKENHGDVENGCPACDCNDT 993
| | : | | | | | | | | : | | : | |

QY 244 GSHSVMLKSGTNILYWRRTGILMGSKAVKPVLVKNITIEGVAYTSECFCKPGTFSN--- 300
| | : | | | | | | | | : | | : | |
Db 994 GS-----IGSDCDQ-----VSGQC-NCKQGVFGKQCD 1019
| | : | | | | | | | | : | | : | |

QY 301 --KPGSPN-----CQVCPRNTYSEKGAKECIRCKDDSQFSGSECTERPPCT-----TKD 348
| | : | | | | | | | | : | | : | |
Db 1020 QCRPSYFNFTDAGCQFCHNIY---GSIEDGKC--DQTTGKCECENVEGTMCEKCADG 1073
| | : | | | | | | | | : | | : | |

QY 349 YFQIHT-----PCDEEGKTOIMYKWIIEPK-ICREDLTDAIRLPPSGEKKDCPPCNPGE 400
| | : | | | | | | | | : | | : | |
Db 1074 YFNITSGDCGDCDPTGSEDVSCNLVTGQCVCCKPGVT-----GLK--CDSCCLPNF 1123
| | : | | | | | | | | : | | : | |

QY 401 Y---NNGSSSCHPCP-PGTFSDGTKECRPCPCAGTEPALGFEYKWNVLPGNMKTSCFNVG 456
| | : | | | | | | | | : | | : | |
Db 1124 YGLTSECTECEPCPAPGVQVCDPIDGSCVCPNPNT-----VGEMCENCTT-- 1167
| | : | | | | | | | | : | | : | |

QY 457 NSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITFVFE 516
| | : | | | | | | | | : | | : | |
Db 1168 -----NAW-----DYHPLN-----GCK----- 1179
| | : | | | | | | | | : | | : | |

QY 517 TLCADCVLYFMVDINRKSTNVVSWGTTKEKQAY-----THIFKNATFTTWFQ 568
| | : | | | | | | | | : | | : | |
Db 1180 -LC--DC-----SDIGSDGGMCTFTGQCKCKAAYVGLKCDLCTGHGFNFPT----- 1223
| | : | | | | | | | | : | | : | |

QY 569 RTNOGDNRFRFINDMKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGHIYEKE 628
| | : | | | | | | | | : | | : | |
Db 1224 -----GCNAAAGTDPLOCKDQGCL--- 1245
| | : | | | | | | | | : | | : | |
```

```
QY 629 TNQCKECPDYLTIHQVYKAEACIPCGPG-----SKNQDHSVCYSDCF----- 673
| | : | | | | | | | | : | | : | |
Db 1246 CNEIGECP-----CKKNVHGFK-CDQCGEGTFLDSSNLKGGCTECF--CFNRTSNCEQSD 1297
| | : | | | | | | | | : | | : | |

QY 674 -----FYHEKEN-QILHYDFSNLSS-----VGSLMNGPS 701
| | : | | | | | | | | : | | : | |
Db 1298 LVWQMYAEDRRAVFQEPWEFYTKKHNNILLREKPSHFNSYPTDATPLYWPLPSTMLGDR 1357
| | : | | | | | | | | : | | : | |

QY 702 FTS-----KGTKYF-HFFNISLCGHEGKKMALCTNNITDFIVK 738
| | : | | | | | | | | : | | : | |
Db 1358 TASYNGFLRFKIWNEDNRRGLHGIRPDQOQYFRHFPQVVIIFGNRIELEHPIHMEINDDGIY 1417
| | : | | | | | | | | : | | : | |

QY 739 EI-----VAGSDDYT-----NLVGAFCVQSTIIPSESKGFRAALSSQSIL 779
| | : | | | | | | | | : | | : | |
Db 1418 KIRLHESEWRVRHSPELTLTRKOMMVALQDTQGIYIRGTYPARGDAINIQEVSLDVAV 1477
| | : | | | | | | | | : | | : | |

QY 780 ADTFIGVTVETTLKNINIKEDM-----FPVPTSQIPDVHFPYKSS----- 819
| | : | | | | | | | | : | | : | |
Db 1478 PESKIVAGLSTT-KAIGVEKCLGCPQGYTGLSCQNPEVGYRKKHREYLNQADDIALIGW 1536
| | : | | | | | | | | : | | : | |

QY 820 TATTSCINGRSTAVKMRCPNPKSGAGVISVPSKCPAGT-----CDGCTFYFLWESAE---- 871
| | : | | | | | | | | : | | : | |
Db 1537 SEPCSC-HGHS---QTCNPD-----TSVCTDCEHNTFGDCEHCLPGYIGDAREGGAN 1585
| | : | | | | | | | | : | | : | |

QY 872 -----ACPL-----CTEHDFFELEGACKRGFQETLYVWNEPKWCIKG-ISLPEKK 915
| | : | | | | | | | | : | | : | |
Db 1586 ACTKCACPLVENSFSDSCVAVDHGRGYVCNSCKPG-----YTGQYCETCVAGYYGDPQHI 1640
| | : | | | | | | | | : | | : | |

QY 916 LATCEIVD 923
| | : | | | | | | | | : | | : | |
Db 1641 GGTCSPCD 1648
| | : | | | | | | | | : | | : | |

RESULT 15
Q9TZR4
ID Q9TZR4 PRELIMINARY; PRT; 3102 AA.
AC Q9TZR4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update).
DE Laminin alpha chain.
GN LAMAL/2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Zhu X., Kao G., Joh K., Hall D.H., Wadsworth V.K., Hutter H.,
RA Vogel B.E., Huang C., Yochem J., Horl K., Hedgecock E.M.,
RA Wadsworth W.G.;
RT "Expression, function and evolution of laminin alpha chains.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF074902; AAC26793.1; -.
DR HSSP; P02468; 1TLE.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR001886; LamNT.
DR Pfam; PF00052; laminin_B; 2.
DR Pfam; PF00053; laminin_EGF; 17.
DR Pfam; PF00054; laminin_G; 3.
DR Pfam; PF00055; laminin_Nterm; 1.
DR PRINTS; PR00011; EGFLAMININ.
DR ProDom; PD002082; LamNT; 1.
DR ProDom; PD003031; Laminin_B; 2.
DR SMART; SM00180; EGF_Lam; 14.
DR SMART; SM00001; EGF_like; 2.
DR SMART; SM00281; LamB; 2.
DR SMART; SM00282; LamG; 4.
```







GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:31:24 ; Search time 17.5467 Seconds  
(without alignments)  
5626.691 Million cell updates/sec

Title: US-10-073-333A-4  
Perfect score: 1027  
Sequence: 1 MLFRARGPVRGRGWGRPAEA.....KEKEDHFESVQLKTSRSPNI 1027

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	8	0.8	100	1	S27016	defensin alpha-6 p
2	8	0.8	148	2	E84239	hypothetical prote
3	8	0.8	184	1	QOCVLG	BL1 protein - toma
4	8	0.8	228	2	S65426	pyruvate decarboxy
5	8	0.8	228	2	S65425	pyruvate decarboxy
6	8	0.8	239	2	G87265	conserved hypothet
7	8	0.8	268	2	AB1424	E. coli Rpir trans
8	8	0.8	268	2	AH1797	E. coli Rpir trans
9	8	0.8	271	2	S12783	OX40 antigen precu
10	8	0.8	272	2	I48700	gene ox40 protein
11	8	0.8	281	2	T33466	hypothetical prote
12	8	0.8	286	2	A82929	ATP synthase gamma
13	8	0.8	405	2	S65471	pyruvate decarboxy
14	8	0.8	417	2	S57820	pyruvate decarboxy
15	8	0.8	428	2	AF0241	probable coenzyme
16	8	0.8	444	2	T01548	protein kinase hom
17	8	0.8	463	2	T28748	hypothetical prote
18	8	0.8	585	2	T03252	pyruvate decarboxy
19	8	0.8	603	2	T03295	pyruvate decarboxy
20	8	0.8	622	2	A36915	fructanase - Bacte
21	8	0.8	671	2	T10755	kinesin-related pr
22	8	0.8	742	2	AE2357	hypothetical prote
23	8	0.8	881	2	T01269	serine/threonine-s
24	8	0.8	1206	2	B87247	probable conserved
25	8	0.8	1578	2	AD1512	peptidoglycan boun
26	8	0.8	1582	2	AC1153	adhesin homolog lm
27	8	0.8	3102	2	T43291	laminin alpha chai
28	7	0.7	94	2	A64863	hypothetical prote
29	7	0.7	100	1	B43719	urease (EC 3.5.1.5

30	7	0.7	100	1	S08478	urease (EC 3.5.1.5
31	7	0.7	100	1	TNLJSI	trans-activating t
32	7	0.7	100	2	B47090	urease (EC 3.5.1.5
33	7	0.7	100	2	A43998	hypothetical prote
34	7	0.7	109	2	S42599	hypothetical prote
35	7	0.7	122	2	T28199	hypothetical prote
36	7	0.7	134	2	C84023	hypothetical prote
37	7	0.7	134	2	T36365	proline-rich prote
38	7	0.7	135	2	PC2269	cytochrome P450 pr
39	7	0.7	141	2	A27482	vasotocin / neurop
40	7	0.7	147	2	E70746	probable mmpS2 pro
41	7	0.7	157	2	AG3560	transcription regu
42	7	0.7	159	2	B29879	vasotocin / neurop
43	7	0.7	173	2	AB3450	invasion protein b
44	7	0.7	179	2	H64472	hypothetical prote
45	7	0.7	180	2	T49530	related to glycine

ALIGNMENTS

RESULT 1  
S27016  
defensin alpha-6 precursor - human  
N:Alternate names: Paneth cell-specific alpha-defensin 6  
C:Species: Homo sapiens (man)  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: S27016  
R:Jones, D.E.; Bevins, C.L.  
FEBS Lett. 315, 187-192, 1993  
A:Title: Defensin-6 mRNA in human Paneth cells: implications for antimicrobial peptid  
A:Reference number: S27016; MUID:93114459; PMID:8417977  
A:Accession: S27016  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-100 <JON>  
A:Cross-references: GB:M98331; NID:g181546; PIDN:AAB59357.1; PID:g181547  
C:Genetics:  
A:Gene: GDB:DEFA6; DEF6; HD-6  
A:Cross-references: GDB:136838; OMIM:600471  
A:Map position: 8pter-8p21  
C:Superfamily: mammalian defensin  
C:Keywords: antibiotic; homodimer  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-70/Domain: propeptide #status predicted <PRO>  
F:71-100/Product: defensin alpha-6 #status predicted <MAT>  
F:72-99,74-88,78-98/Disulfide bonds: #status predicted  
Query Match 0.8%; Score 8; DB 1; Length 100;  
Best Local Similarity 100.0%; Pred. No. 4.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 936 TAVLLVAL 943  
| | | | | | | |  
Db 8 TAVLLVAL 15  
RESULT 2  
E84239  
hypothetical protein Vng0825c [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: E84239  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky  
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: E84239  
A:Status: preliminary  
A:Molecule type: DNA

A;Residues: 1-148 <STO>  
A;Cross-references: GB:AE004437; NID:gl0580395; PIDN:AAG19281.1; GSPDB:GN00138  
C;Genetics:  
A;Gene: VNG0825C

Query Match 0.8%; Score 8; DB 2; Length 148;  
Best Local Similarity 100.0%; Pred. No. 6.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 TVVGPSDS 162  
Db 5 TVVGPSDS 12

RESULT 3  
QQCVLG  
BL1 protein - tomato golden mosaic virus  
C;Species: tomato golden mosaic virus  
A;Note: host Nicotiana sp. (tobacco)  
C;Date: 28-Aug-1985 #sequence\_revision 28-Aug-1985 #text\_change 08-Apr-1994  
C;Accession: A04169  
R;Hamilton, W.D.O.; Stein, V.E.; Coutts, R.H.A.; Buck, K.W.  
EMBO J. 3, 2197-2205, 1984  
A;Title: Complete nucleotide sequence of the infectious cloned DNA components of tomato  
A;Reference number: A04163  
A;Accession: A04169  
A;Molecule type: DNA  
A;Residues: 1-184 <HAM>  
C;Comment: The genome consists of two circular, single-stranded DNA components, DNA A and  
C;Genetics:  
A;Map position: segment B  
C;Superfamily: tomato golden mosaic virus BL1 protein

Query Match 0.8%; Score 8; DB 1; Length 184;  
Best Local Similarity 100.0%; Pred. No. 7.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 NYIESNRD 184  
Db 13 NYIESNRD 20

RESULT 4  
S65426  
pyruvate decarboxylase (EC 4.1.1.1) - fava bean (fragment)  
C;Species: Vicia faba (fava bean)  
C;Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 15-Oct-1999  
C;Accession: S65426  
R;Muecke, U.; Wohlfarth, T.; Fiedler, U.; Baeumlein, H.; Ruecknagel, K.P.; Koenig, S.  
Eur. J. Biochem. 237, 373-382, 1996  
A;Title: Pyruvate decarboxylase from Pisum sativum. Properties, nucleotide and amino acid  
A;Reference number: S65423; MUID:96215432; PMID:8647075  
A;Accession: S65426  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-228 <MUE>  
C;Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain  
C;Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 0.8%; Score 8; DB 2; Length 228;  
Best Local Similarity 100.0%; Pred. No. 9.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLY 276  
Db 50 KAVKPVLY 57

RESULT 5  
S65425  
pyruvate decarboxylase (EC 4.1.1.1) (clone PDC3) - garden pea (fragment)  
C;Species: Pisum sativum (garden pea)

C;Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 15-Oct-1999  
C;Accession: S65425  
R;Muecke, U.; Wohlfarth, T.; Fiedler, U.; Baeumlein, H.; Ruecknagel, K.P.; Koenig, S.  
Eur. J. Biochem. 237, 373-382, 1996  
A;Title: Pyruvate decarboxylase from Pisum sativum. Properties, nucleotide and amino  
A;Reference number: S65423; MUID:96215432; PMID:8647075  
A;Accession: S65425  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-228 <MUE>  
C;Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain  
C;Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 0.8%; Score 8; DB 2; Length 228;  
Best Local Similarity 100.0%; Pred. No. 9.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLY 276  
Db 50 KAVKPVLY 57

RESULT 6  
G87265  
conserved hypothetical protein CC0136 [imported] - Caulobacter crescentus  
C;Species: Caulobacter crescentus  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C;Accession: G87265  
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A;Title: Complete Genome Sequence of Caulobacter crescentus.  
A;Reference number: A87249; MUID:21173698; PMID:11259647  
A;Accession: G87265  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-239 <STO>  
A;Cross-references: GB:AE005673; NID:gl3421247; PIDN:AAK22123.1; GSPDB:GN00148  
C;Genetics:  
A;Gene: CC0136

Query Match 0.8%; Score 8; DB 2; Length 239;  
Best Local Similarity 100.0%; Pred. No. 9.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 AEAPRRGR 25  
Db 176 AEAPRRGR 183

RESULT 7  
AB1424  
E. coli RpiR transcription regulator homolog lmo2795 [imported] - Listeria monocytoge  
C;Species: Listeria monocytogenes  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AB1424  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AB1424  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-268 <GLA>  
A;Cross-references: GB:NC\_003210; PIDN:CAD01008.1; PID:gl6412295; GSPDB:GN00177  
A;Experimental source: strain EGD-e  
C;Genetics:  
A;Gene: lmo2795



Query Match 0.8%; Score 8; DB 2; Length 268;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 792 LKNINIKE 799  
Db 32 LKNINIKE 39  
|||||

RESULT 8  
AH1797  
E. coli RpiR transcription regulator homolog lin2927 [imported] - Listeria innocua (strain  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AH1797  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AH1797  
A;Status: preliminary.  
A;Molecule type: DNA  
A;Residues: 1-268 <GLA>  
A;Cross-references: GB:AL592022; PIDN:CAC98152.1; PID:g16415468; GSPDB:GN00178  
A;Experimental source: strain Clip11262  
C;Genetics:  
A;Gene: lin2927

Query Match 0.8%; Score 8; DB 2; Length 268;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 792 LKNINIKE 799  
Db 32 LKNINIKE 39  
|||||

RESULT 9  
S12783  
OX40 antigen precursor - rat  
N;Alternate names: nerve growth factor receptor homolog  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 05-Nov-1999  
C;Accession: S12783; S08036  
R;Mallett, S.; Fossum, S.; Barclay, A.N.  
EMBO J. 9, 1063-1068, 1990  
A;Title: Characterization of the MRC OX40 antigen of activated CD4 positive T lymphocyte  
A;Reference number: S12783; MUID:90214614; PMID:2157591  
A;Accession: S12783  
A;Molecule type: mRNA  
A;Residues: 1-271 <MAL>  
A;Cross-references: EMBL:X17037; NID:g57830; PIDN:CAA34897.1; PID:g57831  
C;Superfamily: CD27 antigen; NGF receptor repeat homology  
C;Keywords: growth factor receptor; transmembrane protein  
F;1-19/Domain: signal sequence #status predicted <SIG>  
F;20-271/Product: OX40 antigen #status predicted <MAT>  
F;211-235/Domain: transmembrane #status predicted <TM>

Query Match 0.8%; Score 8; DB 2; Length 271;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623  
Db 122 CVPCPPGH 129  
|||||

RESULT 10

I48700  
gene ox40 protein - mouse  
N;Alternate names: OX40 antigen  
C;Species: Mus musculus (house mouse)  
C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 11-Jan-2000  
C;Accession: I48700; I48334; S34377  
R;Calderhead, D.M.; Buhlmann, J.E.; van den Eertwegh, A.J.; Claassen, E.; Noelle, R.J  
J. Immunol. 151, 5261-5271, 1993  
A;Title: Cloning of mouse Ox40: a T cell activation marker that may mediate T-B cell  
A;Reference number: I48700; MUID:94044750; PMID:8228223  
A;Accession: I48700  
A;Status: translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-272 <RES>  
A;Cross-references: EMBL:Z21674; NID:g312827; PIDN:CAA79772.1; PID:g312828  
R;Birkeland, M.L.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.; Barclay, A.N.  
Eur. J. Immunol. 25, 926-930, 1995  
A;Title: Gene structure and chromosomal localization of the mouse homologue of rat OX  
A;Reference number: I48334; MUID:95255413; PMID:7737295  
A;Accession: I48334  
A;Status: translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-14 'G', 16-272 <RE2>  
A;Cross-references: EMBL:X85214; NID:g732818; PIDN:CAA59476.1; PID:g732819  
C;Genetics:  
A;Gene: ox40  
A;Introns: 45/1; 86/1; 122/1; 144/2; 210/1; 250/1  
C;Superfamily: CD27 antigen; NGF receptor repeat homology

Query Match 0.8%; Score 8; DB 2; Length 272;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623  
Db 123 CVPCPPGH 130  
|||||

RESULT 11  
T33466  
hypothetical protein C43H8.3 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T33466  
R;Tin-Wollam, A.M.  
submitted to the EMBL Data Library, October 1998  
A;Description: The sequence of C. elegans cosmid C43H8.  
A;Reference number: Z21350  
A;Accession: T33466  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-281 <TIN>  
A;Cross-references: EMBL:AF098499; PIDN:AAC67398.1; GSPDB:GN00019; CESP:C43H8.3  
A;Experimental source: strain Bristol N2; clone C43H8  
C;Genetics:  
A;Gene: CESP:C43H8.3  
A;Map position: 1  
A;Introns: 58/1; 114/2; 242/3

Query Match 0.8%; Score 8; DB 2; Length 281;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 NITDFTVK 738  
Db 111 NITDFTVK 118  
|||||

RESULT 12  
A82929  
ATP synthase gamma chain UUI30 [imported] - Ureaplasma urealyticum  
C;Species: Ureaplasma urealyticum  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000

C:Accession: A82929  
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.  
submitted to GenBank, February 2000  
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min  
A:Reference number: A82870  
A:Accession: A82929  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-286 <GLA>  
A:Cross-references: GB:AF002114; GB:AF222894; NID:g6899086; PIDN:AAF30536.1; GSPDB:GN001  
A:Experimental source: serovar 3; biovar 1  
C:Genetics:  
A:Gene: atpG; UUI30  
A:Genetic code: SGC3  
C:Superfamily: H<sup>+</sup>-transporting ATP synthase gamma chain

Query Match 0.8%; Score 8; DB 2; Length 286;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 EIVAGSDD 746  
|||||||  
Db 279 EIVAGSDD 286

RESULT 13  
S65471  
pyruvate decarboxylase (EC 4.1.1.1) (clone PDC2) - garden pea (fragment)  
C:Species: Pisum sativum (garden pea)  
C:Date: 19-Mar-1997 #sequence\_revision 06-Jun-1997 #text\_change 20-Jun-2000  
C:Accession: S65471; S65424  
R:Muecke, U.; Wohlfarth, T.; Fiedler, U.; Baeumlein, H.; Ruecknagel, K.P.; Koenig, S.  
Eur. J. Biochem. 237, 373-382, 1996  
A:Title: Pyruvate decarboxylase from Pisum sativum. Properties, nucleotide and amino aci  
A:Reference number: S65423; MUID:96215432; PMID:8647075  
A:Accession: S65471  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-405 <MUE>  
A:Cross-references: EMBL:Z66544; NID:g1177604; PIDN:CAA91445.1; PID:g1177605  
A:Accession: S65424  
A:Molecule type: protein  
A:Residues: 1-8;146-153;209-218;350-362;386-393 <MUW>  
C:Genetics:  
A:Gene: pdc  
C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain  
C:Keywords: carbon-carbon lyase; carboxy-lyase  
F:1-405/Product: pyruvate decarboxylase #status experimental <MAT>  
F:271-317/Domain: thiamin pyrophosphate-binding domain homology <TPB>

Query Match 0.8%; Score 8; DB 2; Length 405;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLV 276  
|||||||  
Db 50 KAVKPVLV 57

RESULT 14  
S57820  
pyruvate decarboxylase (EC 4.1.1.1) 1 - common tobacco (fragment)  
C:Species: Nicotiana tabacum (common tobacco)  
C:Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 15-Oct-1999  
C:Accession: S57820  
R:Bucher, M.; Brander, K.A.; Sbicego, S.; Mandel, T.; Kuhlmeier, C.  
Plant Mol. Biol. 28, 739-750, 1995  
A:Title: Aerobic fermentation in tobacco pollen.  
A:Reference number: S57819; MUID:95375236; PMID:7647304  
A:Accession: S57820  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-417 <BUC>

A:Cross-references: EMBL:X81854  
C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domai  
C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 0.8%; Score 8; DB 2; Length 417;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLV 276  
|||||||  
Db 155 KAVKPVLV 162

RESULT 15  
AF0241  
probable coenzyme synthetase YPO1981 [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
C:Accession: AF0241  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Burrel  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AF0241  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-428 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC90794.1; PID:g15979994; GSPDB:GN00175  
C:Genetics:  
A:Gene: YPO1981

Query Match 0.8%; Score 8; DB 2; Length 428;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 604 ACALGSEQ 611  
|||||||  
Db 84 ACALGSEQ 91

Search completed: May 12, 2003, 13:41:40  
Job time : 19.5467 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:23:58 ; Search time 13.9342 Seconds  
(without alignments)  
3056.959 Million cell updates/sec

Title: US-10-073-333A-4  
Perfect score: 1027  
Sequence: 1 MLFRARGPVRRGGRPAEA.....KEKEDHFESVQLKTSRSPNI 1027

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	8	0.8	100	1	DEF6_HUMAN	Q01524 homo sapien
2	8	0.8	271	1	TNR4_RAT	P15725 rattus norv
3	8	0.8	272	1	TNR4_MOUSE	P47741 mus musculu
4	8	0.8	293	1	VBL1_TGMV	P03566 tomato gold
5	8	0.8	300	1	TR6B_HUMAN	O95407 homo sapien
6	8	0.8	405	1	DCP2_PEA	P51851 pisum sativ
7	8	0.8	418	1	DCP1_TOBAC	P51845 nicotiana t
8	8	0.8	430	1	TRLT_HUMAN	Q96924 homo sapien
9	8	0.8	430	1	TRLT_MACFA	Q9n092 macaca fasc
10	8	0.8	585	1	DCP3_ORYSA	P51849 oryza sativ
11	8	0.8	603	1	DCP2_ORYSA	P51848 oryza sativ
12	8	0.8	671	1	KRP2_RAT	Q62909 rattus norv
13	8	0.8	718	1	MCAK_CRIGR	P70096 cricetulus
14	7	0.7	100	1	TAT_SIVCZ	P17285 chimpanzee
15	7	0.7	100	1	URE3_ECOLI	Q03282 escherichia
16	7	0.7	100	1	URE3_PROMI	P17088 proteus mir
17	7	0.7	100	1	URE3_PROVU	P16124 proteus vul
18	7	0.7	101	1	RK24_GUITH	O46905 guillardia
19	7	0.7	134	1	CRB2_BACHD	Q9k819 bacillus ha
20	7	0.7	141	1	NEUV_RANES	P11858 rana esculie
21	7	0.7	145	1	MMS2_MYCTU	Q11170 mycobacteri
22	7	0.7	159	1	NEUV_BUFJA	P08163 bufo japoni
23	7	0.7	179	1	YCBQ_ECOLI	P75855 escherichia
24	7	0.7	179	1	YD85_METJA	Q58780 methanococc
25	7	0.7	200	1	RS8A_SCHPO	O14049 schizosacch
26	7	0.7	200	1	RS8B_SCHPO	Q9p7b2 schizosacch
27	7	0.7	203	1	Y296_METJA	Q57744 methanococc
28	7	0.7	206	1	Y930_HAEIN	P44077 haemophilus
29	7	0.7	208	1	YM16_YEAST	P40206 saccharomyc
30	7	0.7	208	1	YQED_BACSU	P54449 bacillus su
31	7	0.7	259	1	SP07_YEAST	P18410 saccharomyc
32	7	0.7	270	1	PYRF_CANAL	P13649 candida alb
33	7	0.7	270	1	PYRF_CANDU	Q9c150 candida dub

RESULT 1						
DEF6_HUMAN						
ID	DEF6_HUMAN	STANDARD;	PRT;	100	AA.	
AC	Q01524;					
DT	01-JUL-1993 (Rel. 26, Created)					P21188 xenopus lae
DT	01-JUL-1993 (Rel. 26, Last sequence update)					Q92956 homo sapien
DT	15-JUN-2002 (Rel. 41, Last annotation update)					Q9hc21 homo sapien
DE	Defensin 6 precursor (Defensin, alpha 6).					Q9wy63 thermotoga
GN	DEFA6 OR DEF6.					P53425 mycobacteri
OS	Homo sapiens (Human).					O14904 homo sapien
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					O14303 schizosacch
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					Q94175 caenorhabdi
OX	NCBI_TaxID=9606;					P54465 bacillus su
RN	[1]					P05432 rattus norv
RP	SEQUENCE FROM N.A.					P28003 saccharomyc
RC	TISSUE=Intestine;					Q11178 caenorhabdi
RX	MEDLINE=93114459; PubMed=8417977;					
RA	Jones D.E., Bevins C.L.;					
RT	"Defensin-6 mRNA in human Paneth cells: implications for antimicrobial peptides in host defense of the human bowel.";					
RL	FEBS Lett. 315:187-192(1993).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=96223969; PubMed=8626737;					
RA	Mallow E.B., Harris A., Salzman N., Russell J.P.,					
RA	Deberardinis R.J., Ruchelli E., Bevins C.L.;					
RT	"Human enteric defensins. Gene structure and developmental expression.";					
RL	J. Biol. Chem. 271:4038-4045(1996).					
CC	-!- FUNCTION: HAS ANTIMICROBIAL ACTIVITY.					
CC	-!- TISSUE SPECIFICITY: PANETH CELLS OF THE SMALL INTESTINE.					
CC	-!- SIMILARITY: BELONGS TO THE CORTICOSTATIN/DEFENSIN FAMILY.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).					
CC	-----					
DR	EMBL; M98331; AAB59357.1; -.					
DR	EMBL; U33317; AAC50382.1; ALT_SEQ.					
DR	PIR; S27016; S27016.					
DR	Genew; HGNC:2765; DEFA6.					
DR	MIM; 600471; -.					
DR	InterPro; IPR002366; Defensin_alpha.					
DR	InterPro; IPR001271; Defensin_mammal.					
DR	Pfam; PF00323; defensins; 1.					
DR	Pfam; PF00879; Defensin_propep; 1.					
DR	SMART; SM00048; DEFSN; 1.					
DR	PROSITE; PS00269; DEFENSIN; 1.					
KW	Defensin; Antibiotic; Fungicide; Signal.					
FT	SIGNAL 1 19 POTENTIAL.					
FT	PROPEP 20 265 POTENTIAL.					
FT	CHAIN 266 100 DEFENSIN 6.					

ALIGNMENTS



```
FT DISULFID 72 99 BY SIMILARITY.
FT DISULFID 74 88 BY SIMILARITY.
FT DISULFID 78 98 BY SIMILARITY.
SQ SEQUENCE 100 AA; 10975 MW; EDF7E033DDCE2D5 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 TAVLLVAL 943
Db 8 TAVLLVAL 15

RESULT 2
TNR4_RAT
ID TNR4_RAT STANDARD; PRT; 271 AA.
AC P15725;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor necrosis factor receptor superfamily member 4 precursor (OX40L
DE receptor) (OX40 antigen) (MRC OX40).
GN TNFRSF4 OR TXGPIL OR OX40.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=T-cell;
RX MEDLINE=90214614; PubMed=2157591;
RA Mallett S., Fossum S., Barclay A.N.;
RT "Characterization of the MRC OX40 antigen of activated CD4 positive T
RT lymphocytes -- a molecule related to nerve growth factor receptor.";
RL EMBO J. 9:1063-1068(1990).
CC -!- FUNCTION: Receptor for TNFSF4/OX40L/GP34.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: ACTIVATED T-CELLS.
CC -!- SIMILARITY: CONTAINS 4 TNFR-CYS REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X17037; CAA34897.1; -.
CC PIR; S08036; S08036.
CC PIR; S12783; S12783.
CC HSSP; O14763; 1D4V.
CC InterPro; IPR001368; TNFR_c6.
CC Pfam; PF00020; TNFR_c6; 3.
CC ProDom; PD000771; TNFR_c6; 1.
CC SMART; SM00208; TNFR; 3.
CC PROSITE; PS00652; TNFR_NGFR_1; 2.
CC PROSITE; PS50050; TNFR_NGFR_2; 2.
CC Receptor; Antigen; Transmembrane; Glycoprotein; Repeat; Signal.
KW SIGNAL 1 19
FT CHAIN 20 271
FT FT
FT FT
FT DOMAIN 20 210
FT TRANSMEM 211 235
FT DOMAIN 236 271
FT REPEAT 25 60
FT REPEAT 61 102
FT REPEAT 103 123
FT REPEAT 124 164
FT DISULFID 26 37
FT DISULFID 38 51
FT DISULFID 41 59
```

```
FT DISULFID 62 76 BY SIMILARITY.
FT DISULFID 79 94 BY SIMILARITY.
FT DISULFID 82 102 BY SIMILARITY.
FT DISULFID 104 122 BY SIMILARITY.
FT DISULFID 125 138 BY SIMILARITY.
FT DISULFID 144 163 BY SIMILARITY.
FT CARBOHYD 143 143 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 271 AA; 29895 MW; C06465136B16E821 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 271;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623
Db 122 CVPCPPGH 129

RESULT 3
TNR4_MOUSE
ID TNR4_MOUSE STANDARD; PRT; 272 AA.
AC P47741;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor necrosis factor receptor superfamily member 4 precursor (OX40L
DE receptor) (OX40 antigen).
GN TNFRSF4 OR TXGP1 OR OX40.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX MEDLINE=94044750; PubMed=8228223;
RA Calderhead D.M., Buhlmann J.E., van den Eertwegh A.J.,
RA Claassen E., Noelle R.J., Fell H.;
RT "Cloning of mouse OX40: a T cell activation marker that may mediate
RT T-B cell interactions.";
RL J. Immunol. 151:5261-5271(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95255413; PubMed=7737295;
RA Birkeland M.L., Copeland N.G., Gilbert D.J., Jenkins N.A.,
RA Barclay A.N.;
RT "Gene structure and chromosomal localization of the mouse homologue
RT of rat OX40 protein.";
RL Eur. J. Immunol. 25:926-930(1995).
CC -!- FUNCTION: Receptor for TNFSF4/OX40L/GP34.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: CONTAINS 4 TNFR-CYS REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z21674; CAA79772.1; -.
CC EMBL; X85214; CAA59476.1; -.
CC HSSP; O14763; 1D0G.
CC MGD; MGI:104512; Tnfrsf4.
CC InterPro; IPR001368; TNFR_c6.
CC Pfam; PF00020; TNFR_c6; 3.
CC ProDom; PD000771; TNFR_c6; 1.
CC SMART; SM00208; TNFR; 3.
CC PROSITE; PS00652; TNFR_NGFR_1; 2.
CC PROSITE; PS50050; TNFR_NGFR_2; 2.
CC Receptor; Antigen; Transmembrane; Glycoprotein; Repeat; Signal.
KW SIGNAL 1 19
FT CHAIN 20 271
```

```
FT CHAIN 20 272 TUMOR NECROSIS FACTOR RECEPTOR
FT SUPERFAMILY MEMBER 4.
FT DOMAIN 20 211 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 212 236. POTENTIAL.
FT DOMAIN 237 272 CYTOPLASMIC (POTENTIAL).
FT REPEAT 26 61 TNFR-CYS 1.
FT REPEAT 62 103 TNFR-CYS 2.
FT REPEAT 104 124 TNFR-CYS 3 (INCOMPLETE).
FT REPEAT 125 165 TNFR-CYS 4.
FT DISULFID 27 38 BY SIMILARITY.
FT DISULFID 39 52 BY SIMILARITY.
FT DISULFID 42 60 BY SIMILARITY.
FT DISULFID 63 77 BY SIMILARITY.
FT DISULFID 80 95 BY SIMILARITY.
FT DISULFID 83 103 BY SIMILARITY.
FT DISULFID 105 123 BY SIMILARITY.
FT DISULFID 126 139 BY SIMILARITY.
FT DISULFID 145 164 BY SIMILARITY.
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 15 15 A -> G (IN REF. 2).
SQ SEQUENCE 272 AA; 30153 MW; 06E7BB4156F0D08E CRC64;

Query Match 0.8%; Score 8; DB 1; Length 272;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623
DB 123 CVPCPPGH 130

RESULT 4
VBL1_TGMV STANDARD; PRT; 293 AA.
AC P03566;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE BL1 protein.
GN BCL1.
OS Tomato golden mosaic virus (TGMV).
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX NCBI_TaxID=10831;
RN [1]
RP SEQUENCE FROM N.A.
RA von Arnim A.G., Stanley J.;
RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP PRELIMINARY SEQUENCE FROM N.A.
RA Hamilton W.D.O., Stein V.E., Coutts R.H.A., Buck K.W.;
RT "Complete nucleotide sequence of the infectious cloned DNA components
RT of tomato golden mosaic virus: potential coding regions and regulatory
RT sequences."
RL EMBO J. 3:2197-2205(1984).
CC -!- SIMILARITY: BELONGS TO GEMINIVIRUSES BL1 PROTEIN FAMILY.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 173
CC ONWARD DUE TO A FRAMESHIFT ERROR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M73794; AAA46584.1; -
CC EMBL; K02030; -; NOT_ANNOTATED_CDS.
CC PIR; A04169; QQCVLG.
CC InterPro; IPR000211; Gemini_BL.
CC Pfam; PF00845; Gemini_BL1; 1.
FT CONFLICT 11 11 V -> A (IN REF. 2).
SQ SEQUENCE 293 AA; 32933 MW; E024E3F285C89012 CRC64;
```

```
Query Match 0.8%; Score 8; DB 1; Length 293;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 NYIESNRD 184
DB 13 NYIESNRD 20

RESULT 5
TR6B_HUMAN STANDARD; PRT; 300 AA.
ID TR6B_HUMAN
AC O95407;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor necrosis factor receptor superfamily member 6B precursor (Decoy
DE receptor for Fas ligand) (Decoy receptor 3) (DCR3) (M68).
GN TNFRSF6B OR DCR3 OR TR6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal lung;
RX MEDLINE=99087326; PubMed=9872321;
RA Pitti R.M., Marsters S.A., Lawrence D.A., Roy M., Kischkel F.C.,
RA Dowd P., Huang A., Donahue C.J., Sherwood S.W., Baldwin D.T.,
RA Godowski P.J., Wood W.I., Gurney A.L., Hillan K.J., Cohen R.L.,
RA Goddard A.D., Botstein D., Ashkenazi A.;
RT "Genomic amplification of a decoy receptor for Fas ligand in lung and
RT colon cancer."
RL Nature 396:699-703(1998).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-35.
RC TISSUE=Prostate;
RX MEDLINE=99253915; PubMed=10318773;
RA Yu K.-Y., Kwon B., Ni J., Zhai Y., Ebner R., Kwon B.S.;
RT "A newly identified member of tumor necrosis factor receptor
RT superfamily (TR6) suppresses LIGHT-mediated apoptosis."
RL J. Biol. Chem. 274:13733-13736(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=20122600; PubMed=10655513;
RA Bai C., Connolly B., Metzker M.L., Hilliard C.A., Liu X., Sandig V.,
RA Soderman A., Galloway S.M., Liu Q., Austin C.P., Caskey C.T.;
RT "Overexpression of M68/DCR3 in human gastrointestinal tract tumors
RT independent of gene amplification and its location in a four-gene
RT cluster."
RL Proc. Natl. Acad. Sci. U.S.A. 97:1230-1235(2000).
RN [4]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Decoy receptor for the cytotoxic ligands TNFS14/LIGHT
CC and TNFSF6/FasL. Protects against apoptosis.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Detected in fetal lung, brain and liver.
CC Detected in adult stomach, spinal cord, lymph node, trachea,
CC spleen, colon and lung. Highly expressed in several primary tumors
CC from colon, stomach, rectum, esophagus and in SW480 colon
CC carcinoma cells.
CC -!- SIMILARITY: CONTAINS 4 TNFR-CYS REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF104419; AAD03056.1; -  
DR EMBL; AF134240; AAD29688.1; -  
DR EMBL; AF217796; AAF35244.1; -  
DR EMBL; AF217793; AAF33685.1; -  
DR EMBL; AF217794; AAF33686.1; -  
DR EMBL; AL121845; AAC03668.1; -  
DR EMBL; BC017065; AAH17065.1; -  
DR Genew; HGNC:11921; TNFRSF6B.  
DR MIM; 603361; -  
DR HSSP; O14763; IDOG.  
DR InterPro; IPR001368; TNFR\_c6.  
DR Pfam; PF00020; TNFR\_c6; 4.  
DR ProDom; PD000771; TNFR\_c6; 1.  
DR SMART; SM00208; TNFR; 3.  
DR PROSITE; PS00652; TNFR\_NGFR\_1; 2.  
DR PROSITE; PS50050; TNFR\_NGFR\_2; 2.  
KW Receptor; Apoptosis; Glycoprotein; Repeat; Signal.  
FT SIGNAL 1 29  
FT CHAIN 30 300 TUMOR NECROSIS FACTOR RECEPTOR  
FT SUPERFAMILY MEMBER 6B.  
FT REPEAT 31 70  
FT REPEAT 72 113 TNFR-CYS 1.  
FT REPEAT 115 150 TNFR-CYS 2.  
FT REPEAT 152 193 TNFR-CYS 3.  
FT REPEAT 49 62 TNFR-CYS 4.  
FT DISULFID 52 70 BY SIMILARITY.  
FT DISULFID 73 88 BY SIMILARITY.  
FT DISULFID 91 105 BY SIMILARITY.  
FT DISULFID 95 113 BY SIMILARITY.  
FT DISULFID 115 126 BY SIMILARITY.  
FT DISULFID 132 150 BY SIMILARITY.  
FT DISULFID 153 168 BY SIMILARITY.  
FT DISULFID 174 193 BY SIMILARITY.  
FT CARBOHYD 173 173 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT SEQUENCE 300 AA; 32679 MW; F90AEE33718449AF CRC64;  
Query Match 0.8%; Score 8; DB 1; Length 300;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 410 PCPPGTFS 417  
Db 152 PCPPGTFS 159  
RESULT 6  
DCP2\_PEA STANDARD; PRT; 405 AA.  
AC P51851;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Pyruvate decarboxylase isozyme 2 (EC 4.1.1.1) (PDC) (Fragment).  
GN PDC2.  
OS Pisum sativum (Garden pea).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.  
OX NCBI\_TaxID=3888;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Miko;  
RX MEDLINE=96215432; PubMed=8647075;  
RA Muecke U., Wohlfarth T., Fiedler U., Baumelein H.,  
RA Ruecknagel K.P., Koenig S.;  
RT "Pyruvate decarboxylase from Pisum sativum. Properties, nucleotide

RT and amino acid sequences.";  
RL Eur. J. Biochem. 237:373-382(1996).  
CC -!- CATALYTIC ACTIVITY: A 2-oxo acid = an aldehyde + CO(2).  
CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per  
CC subunit.  
CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z66544; CAA91445.1; -  
DR HSSP; P06672; L2PD.  
DR InterPro; IPR000399; TPP\_enzyme.  
DR Pfam; PF00205; TPP\_enzymes; 1.  
DR Pfam; PF02775; TPP\_enzymes\_C; 1.  
DR PROSITE; PS00187; TPP\_ENZYMES; PARTIAL.  
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;  
KW Multigene family.  
FT NON\_TER 1  
FT SEQUENCE 405 AA; 44078 MW; 44DDCE90B38677FB CRC64;  
Query Match 0.8%; Score 8; DB 1; Length 405;  
Best Local Similarity 100.0%; Pred. No. 6.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 269 KAVKPVLY 276  
Db 50 KAVKPVLY 57  
RESULT 7  
DCP1\_TOBAC STANDARD; PRT; 418 AA.  
ID DCP1\_TOBAC  
AC P51845;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Pyruvate decarboxylase isozyme 1 (EC 4.1.1.1) (PDC) (Fragment).  
GN PDC1.  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Samsun; TISSUE=Leaf;  
RX MEDLINE=95375236; PubMed=7647304;  
RA Bucher M., Brander K., Sbicego S., Mandel T., Kuhlmeier C.;  
RT "Aerobic fermentation in tobacco pollen.";  
RL Plant Mol. Biol. 28:739-750(1995).  
CC -!- CATALYTIC ACTIVITY: A 2-oxo acid = an aldehyde + CO(2).  
CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per  
CC subunit.  
CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).  
CC -!- TISSUE SPECIFICITY: LEAVES.  
CC -!- INDUCTION: ANAEROBICALLY.  
CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----



```
DR EMBL; X81854; CAA57447.1; -
DR HSSP; P06169; LPVD.
DR InterPro; IPR000399; TPP_enzyme.
DR Pfam; PF00205; TPP_enzymes; 1.
DR Pfam; PF02775; TPP_enzymes_C; 1.
DR Pfam; PF02776; TPP_enzymes_N; 1.
DR PROSITE; PS00187; TPP_ENZYMES; PARTIAL.
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;
KW Multigene family.
FT NON_TER 1 1
FT NON_TER 418 418
SQ SEQUENCE 418 AA; 45010 MW; 7C3C33DAA9FFD804 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 418;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 269 KAVKPVLV 276
Db 155 KAVKPVLV 162

RESULT 8
TRLT_HUMAN
ID TRLT_HUMAN STANDARD; PRT; 430 AA.
AC Q969Z4; Q96JU1; Q9BUX7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor necrosis factor receptor superfamily member TNFRSF19L precursor
DE (Receptor expressed in lymphoid tissues).
GN TNFRSF19L OR RELT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF N-TERMINUS.
RC TISSUE=Lymphoma;
RX MEDLINE=21213541; PubMed=11313261;
RA Sica G.L., Zhu G., Tamada K., Liu D., Ni J., Chen L.;
RT "RELt, a new member of the tumor necrosis factor receptor superfamily,
RT is selectively expressed in hematopoietic tissues and activates
RT transcription factor NF-kappaB.";
RL Blood 97:2702-2707(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Retinoblastoma;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon, and Eye;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBBJ databases.
RN [4]
RP SEQUENCE OF 121-430 FROM N.A.
RC TISSUE=Spleen;
RA Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RT spleen.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBBJ databases.
CC -!- FUNCTION: Mediates activation of NF-kappa-B. May play a role in T-
CC cell activation.
CC -!- SUBUNIT: Associates with TRAF1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
```

```
CC -!- TISSUE SPECIFICITY: Highest levels are in spleen, lymph node,
CC thymus, peripheral blood leukocytes, bone marrow and fetal liver.
CC Very low levels in skeletal muscle, testis and colon. Not detected
CC in brain, kidney and pancreas.
CC -!- SIMILARITY: CONTAINS 1 TNFR-CYS REPEAT.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to several
CC frameshifts.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF319553; AAK77356.1; -
CC EMBL; AK027899; BAB55441.1; -
CC EMBL; BC001812; AAH01812.1; -
CC EMBL; BC017279; AAH17279.1; -
CC EMBL; AK074128; BAB84954.1; ALT_FRAME.
CC Genew; HGNC:13764; TNFRSF19L.
CC PROSITE; PS00652; TNFR_NGFR_1; FALSE_NEG.
CC PROSITE; PS50050; TNFR_NGFR_2; FALSE_NEG.
KW Receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 26
FT CHAIN 27 430
FT TUMOR NECROSIS FACTOR RECEPTOR
FT SUPERFAMILY MEMBER TNFRSF19L.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT REPEAT 50 90
FT TNFR-CYS.
FT DISULFID 51 65
FT BY SIMILARITY.
FT DISULFID 71 90
FT BY SIMILARITY.
FT CARBOHYD 149 149
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 122 122
FT D -> S (IN REF. 4).
FT CONFLICT 187 187
FT K -> E (IN REF. 2).
FT CONFLICT 273 273
FT H -> R (IN REF. 2).
FT CONFLICT 379 380
FT DL -> TR (IN REF. 3; AAH01812).
SQ SEQUENCE 430 AA; 46092 MW; 4A5AB9AE32D36101 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 410 PCPPGTFs 417
Db 50 PCPPGTFs 57

RESULT 9
TRLT_MACFA
ID TRLT_MACFA STANDARD; PRT; 430 AA.
AC Q9N092;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor necrosis factor receptor superfamily member TNFRSF19L precursor
DE (Receptor expressed in lymphoid tissues).
GN TNFRSF19L OR RELT.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21458551; PubMed=11574149;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirata M., Suto Y.,
RA Hirai M., Terao K., Suzuki Y., Sugano S., Hashimoto K., Kusuda J.;
RT "Assignment of 118 novel cDNAs of cynomolgus monkey brain to human
RT chromosomes.";
```

RL Gene 275:31-37(2001).  
CC -!- FUNCTION: Mediates activation of NF-kappa-B (By similarity). May  
CC play a role in T-cell activation.  
CC -!- SUBUNIT: Associates with TRAF1 (By similarity).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (probable).  
CC -!- SIMILARITY: CONTAINS 1 TNFR-CYS REPEAT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AB046039; BAB01621.1; -.  
DR InterPro; IPR001368; TNFR\_C6.  
DR PROSITE; PS00652; TNFR\_NGFR\_1; FALSE\_NEG.  
DR PROSITE; PS50050; TNFR\_NGFR\_2; FALSE\_NEG.  
DR SMART; SM00208; TNFR; 1.  
KW Receptor; Transmembrane; Glycoprotein; Signal.  
FT SIGNAL 1 26 BY SIMILARITY.  
FT CHAIN 27 430 TUMOR NECROSIS FACTOR RECEPTOR  
FT SUPERFAMILY MEMBER TNFRSF19L.  
FT DOMAIN 27 162 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 163 183 POTENTIAL.  
FT DOMAIN 184 430 CYTOPLASMIC (POTENTIAL).  
FT REPEAT 50 90 TNFR-CYS.  
FT DISULFID 51 65 BY SIMILARITY.  
FT DISULFID 71 90 BY SIMILARITY.  
FT CARBOHYD 149 149 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 430 AA; 45850 MW; BA8DE92593E1E859 CRC64;  
  
Query Match 0.8%; Score 8; DB 1; Length 430;  
Best Local Similarity 100.0%; Pred. No. 7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 410 PCPPGTF 417  
Db 50 PCPPGTF 57  
IIIIIIII  
  
RESULT 10  
DCP3\_ORYSA STANDARD; PRT; 585 AA.  
ID DCP3\_ORYSA  
AC P51849;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Pyruvate decarboxylase isozyme 3 (EC 4.1.1.1) (PDC).  
GN PDC3.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv Indica-IR54; TISSUE=Callus;  
RX MEDLINE=95148752; PubMed=7846174;  
RA Hossain M.A., McGee J.D., Grover A., Dennis E., Peacock W.J.,  
RA Hodges T.K.;  
RT "Nucleotide sequence of a rice genomic pyruvate decarboxylase gene  
RT that lacks introns: a pseudo-gene?";  
RL Plant Physiol. 106:1697-1698(1994).  
CC -!- CATALYTIC ACTIVITY: A 2-oxo acid = an aldehyde + CO(2).  
CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per  
CC subunit.  
CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U27350; AAA90948.1; -.  
DR EMBL; U38199; AAB40530.1; -.  
DR HSSP; P06169; IPVD.  
DR InterPro; IPR000399; TPP\_enzyme.  
DR Pfam; PF00205; TPP\_enzymes; 1.  
DR Pfam; PF02775; TPP\_enzymes\_C; 1.  
DR PROSITE; PS00187; TPP\_ENZYMES; FALSE\_NEG.  
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;  
KW Multigene family.  
FT ACT\_SITE 89 89 BY SIMILARITY.

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U07338; AAA68289.1; -.  
DR HSSP; P06169; IPVD.  
DR InterPro; IPR000399; TPP\_enzyme.  
DR Pfam; PF00205; TPP\_enzymes; 1.  
DR Pfam; PF02775; TPP\_enzymes\_C; 1.  
DR Pfam; PF02776; TPP\_enzymes\_N; 1.  
DR PROSITE; PS00187; TPP\_ENZYMES; FALSE\_NEG.  
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;  
KW Multigene family.  
FT ACT\_SITE 71 71 BY SIMILARITY.  
SQ SEQUENCE 585 AA; 62456 MW; F26AFA5EC8AAE52 CRC64;  
  
Query Match 0.8%; Score 8; DB 1; Length 585;  
Best Local Similarity 100.0%; Pred. No. 9.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 269 KAVKPVLV 276  
Db 231 KAVKPVLV 238  
IIIIIIII  
  
RESULT 11  
DCP2\_ORYSA STANDARD; PRT; 603 AA.  
ID DCP2\_ORYSA  
AC P51848;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Pyruvate decarboxylase isozyme 2 (EC 4.1.1.1) (PDC).  
GN PDC2.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv Indica-IR54; TISSUE=Callus;  
RA Huq M.E., Hossain M.A., Hodges T.K.;  
RT "Cloning and sequencing of a cDNA encoding pyruvate decarboxylase 2  
RT gene from rice";  
RL (In) Plant Gene Register PGR95-072.  
CC -!- CATALYTIC ACTIVITY: A 2-oxo acid = an aldehyde + CO(2).  
CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per  
CC subunit.  
CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U27350; AAA90948.1; -.  
DR EMBL; U38199; AAB40530.1; -.  
DR HSSP; P06169; IPVD.  
DR InterPro; IPR000399; TPP\_enzyme.  
DR Pfam; PF00205; TPP\_enzymes; 1.  
DR Pfam; PF02775; TPP\_enzymes\_C; 1.  
DR Pfam; PF02776; TPP\_enzymes\_N; 1.  
DR PROSITE; PS00187; TPP\_ENZYMES; FALSE\_NEG.  
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;  
KW Multigene family.  
FT ACT\_SITE 89 89 BY SIMILARITY.





DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE TAT protein (Transactivating regulatory protein).  
GN TAT.  
OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).  
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11723;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90259077; PubMed=2188136;  
RA Huet T., Cheynier R., Meyerhans A., Roelants G., Wain-Hobson S.;  
RT "Genetic organization of a chimpanzee lentivirus related to HIV-1.";  
RL Nature 345:356-359(1990).  
CC -1- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT ACTS BY BINDING TO THE  
CC TRANS-ACTIVATING RESPONSIVE SEQUENCE (TAR) RNA ELEMENT AND  
CC ACTIVATES TRANSCRIPTION INITIATION AND/OR ELONGATION FROM THE LTR  
CC PROMOTER.  
CC -1- SUBUNIT: BINDS CYCLIN T1 (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X52154; CAA36404.1; -.  
DR PIR; S09987; TNLJSI.  
DR HIV; X52154; TAT\$CPZ.  
DR InterPro; IPR001831; HIV\_Tat.  
DR Pfam; PF00539; Tat; 1.  
DR PRINTS; PR00055; HIVTATDOMAIN.  
KW Transcription regulation; Activator; RNA-binding; Nuclear protein;  
KW AIDS.  
SQ SEQUENCE 100 AA; 11209 MW; 1B78830B90EED50E CRC64;  
  
Query Match 0.7%; Score 7; DB 1; Length 100;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 658 GSKNNQD 664  
Db 62 GSKNNQD 68  
  
RESULT 15  
URE3\_ECOLI  
ID URE3\_ECOLI STANDARD; PRT; 100 AA.  
AC Q03282;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Urease gamma subunit (EC 3.5.1.5) (Urea amidohydrolase).  
GN UREA.  
OS Escherichia coli.  
OG Plasmid.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93194816; PubMed=8449894;  
RA D'Orazio S.E., Collins C.M.;  
RT "Characterization of a plasmid-encoded urease gene cluster found in  
RT members of the family Enterobacteriaceae.";  
RL J. Bacteriol. 175:1860-1864(1993).  
CC -1- CATALYTIC ACTIVITY: Urea + H(2)O -> CO(2) + 2 NH(3).  
CC -1- SUBUNIT: (ALPHA, BETA, GAMMA)(3) (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE UREASE GAMMA SUBUNIT FAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L03307; AAA24745.1; -.  
DR PIR; B47090; B47090.  
DR HSSP; P18316; lFWB.  
DR InterPro; IPR002026; Urease\_gamma.  
DR Pfam; PF00547; urease\_gamma; 1.  
DR ProDom; PD002319; Urease\_gamma; 1.  
DR TIGRFAMs; TIGR00193; urease\_gam; 1.  
KW Hydrolase; Plasmid.  
SQ SEQUENCE 100 AA; 10983 MW; FBFD74F4C86B0805 CRC64;  
  
Query Match 0.7%; Score 7; DB 1; Length 100;  
Best Local Similarity 100.0%; Pred. NO. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 977 SCAIMEG 983  
Db 40 SCAIMEG 46  
  
Search completed: May 12, 2003, 13:39:37  
Job time : 16.9342 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.  
OM protein - protein search, using sw model  
Run on: May 12, 2003, 13:30:09 ; Search time 33.5452 Seconds  
(without alignments)  
6308.212 Million cell updates/sec  
Title: US-10-073-333A-4  
Perfect score: 1027  
Sequence: 1 MLFRARGPVRGRGWGRPAEA.....KEKEDHFESVQLKTSRSPNI 1027  
Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0  
Searched: 671580 seqs, 206047115 residues  
Word size : 0  
Total number of hits satisfying chosen parameters: 671580  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Listing first 45 summaries  
Database : SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	342	33.3	Q96DP2 homo sapien
2	21	2.0	Q8R215 mus musculu
3	21	2.0	Q9P2M2 homo sapien
4	9	0.9	Q82444 nicotiana t
5	8	0.8	Q67608 tomato gold
6	8	0.8	Q9hr77 halobacteri
7	8	0.8	Q9dgz5 hepatitis b
8	8	0.8	Q9e6t5 hepatitis b
9	8	0.8	Q9abt4 caulobacter
10	8	0.8	Q926w3 listeria in
11	8	0.8	Q8y3p0 listeria mo
12	8	0.8	Q9tzn3 caenorhabdi
13	8	0.8	Q9prl4 ureaplasma
14	8	0.8	Q8v9i1 squash yell
15	8	0.8	Q50198 mycobacteri
16	8	0.8	Q9tfr5 etheostoma

17	8	0.8	380	8	Q958X6	Q958x6 etheostoma
18	8	0.8	428	16	Q8ZF14	Q8zfl4 yersinia pe
19	8	0.8	444	10	Q23079	Q23079 arabidopsis
20	8	0.8	447	5	Q44593	Q44593 caenorhabdi
21	8	0.8	498	4	Q43267	Q43267 homo sapien
22	8	0.8	550	4	Q9P2Q1	Q9p2q1 homo sapien
23	8	0.8	593	10	Q8RUU6	Q8ruu6 zea mays (m
24	8	0.8	597	10	Q9LGL5	Q9lgl5 oryza sativ
25	8	0.8	605	10	Q8S4W8	Q8s4w8 zea mays (m
26	8	0.8	606	10	Q8S4W9	Q8s4w9 zea mays (m
27	8	0.8	607	10	Q9FFT4	Q9fft4 arabidopsis
28	8	0.8	607	10	Q96536	Q96536 arabidopsis
29	8	0.8	622	2	Q45155	Q45155 bacteroides
30	8	0.8	704	10	Q04434	Q04434 pisum sativ
31	8	0.8	721	11	Q922S8	Q922s8 mus musculu
32	8	0.8	742	16	Q8YNZ6	Q8ynz6 anabaena sp
33	8	0.8	881	10	Q65924	Q65924 arabidopsis
34	8	0.8	1051	12	Q91EU4	Q91eu4 cydia pomon
35	8	0.8	1059	10	Q9LHS4	Q9lhs4 arabidopsis
36	8	0.8	1206	16	Q9CCX9	Q9ccx9 mycobacteri
37	8	0.8	1232	10	Q9LJQ1	Q9ljql arabidopsis
38	8	0.8	1578	16	Q92E25	Q92e25 listeria in
39	8	0.8	1582	16	Q8Y9A5	Q8y9a5 listeria mo
40	8	0.8	2712	16	Q9F3X5	Q9f3x5 pasteurella
41	8	0.8	3102	5	Q9TZR4	Q9tizr4 caenorhabdi
42	8	0.8	3503	5	Q24292	Q24292 drosophila
43	7	0.7	65	6	Q95J37	Q95j37 sus scrofa
44	7	0.7	79	2	Q9AFQ6	Q9afq6 shigella fl
45	7	0.7	84	11	Q8VIN6	Q8vin6 mus musculu

ALIGNMENTS

RESULT 1  
Q96DP2 ID Q96DP2 PRELIMINARY; PRT; 493 AA.  
AC Q96DP2;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE CDNA FLJ31340 fis, clone MESAN1000035, weakly similar to major  
DE surface-labeled trophozoite antigen precursor.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,  
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,  
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,  
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,  
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,  
RA Isogai T.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK055902; BAB71041.1; -  
DR InterPro; IPR001881; EGF\_Ca.  
DR SMART; SM00179; EGF\_CA; 1.  
SQ SEQUENCE 493 AA; 54585 MW; 0AEBB2C19709B211 CRC64;

Query Match 33.3%; Score 342; DB 4; Length 493;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	334	GSSECTERPPCTTKDYFQIHPCDEEGKQIMYKWIPEKICREDLTDAIRLPPSGEKKDC	393
Db	96	GSSECTERPPCTTKDYFQIHPCDEEGKQIMYKWIPEKICREDLTDAIRLPPSGEKKDC	155
QY	394	PCNPGFYNNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCF	453

Db 156 PBCNPGFYNGSSSCHPCPGTFS DGTKECRPCPAGTEPALGF EYKWNVLPGNMKTS CF 215  
QY 454 NVGNSKCDGMNGWEVAGDHQSGAGGSDNDYLILNLHIPGFKPPTSM TGATGSELGRITF 513  
Db 216 NVGNSKCDGMNGWEVAGDHQSGAGGSDNDYLILNLHIPGFKPPTSM TGATGSELGRITF 275  
QY 514 VFETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTTWFQRTNOG 573  
Db 276 VFETLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTTWFQRTNOG 335  
QY 574 QDNRRFINDMVKIYSITATNAV DGVASSCRACALGSEQSGSSCVPCPPGHYIEKETNOCK 633  
Db 336 QDNRRFINDMVKIYSITATNAV DGVASSCRACALGSEQSGSSCVPCPPGHYIEKETNOCK 395  
QY 634 ECPDPTYLSIHQVYGKEACIPCGPGSKNNQDHSV CYSDCFEY 675  
Db 396 ECPDPTYLSIHQVYGKEACIPCGPGSKNNQDHSV CYSDCFEY 437  
RESULT 2  
Q8R215  
ID Q8R215 PRELIMINARY; PRT; 300 AA.  
AC Q8R215;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical 32.6 kDa protein (Fragment).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=COLON;  
RA Strausberg R.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC022655; AAH22655.1; -.  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 300 AA; 32642 MW; 6455109C054B6CB4 CRC64;  
Query Match 2.0%; Score 21; DB 11; Length 300;  
Best Local Similarity 100.0%; Pred. No. 4e-13;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 943 LTCYFWKKNQKLEYKYSKLV M 963  
Db 213 LTCYFWKKNQKLEYKYSKLV M 233  
RESULT 3  
Q9P2M2  
ID Q9P2M2 PRELIMINARY; PRT; 580 AA.  
AC Q9P2M2;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE KIAA1324 protein (Fragment).  
GN KIAA1324.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=20181126; PubMed=10718198;  
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes.XVI.  
RT The complete sequences of 150 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 7:65-73(2000).

DR EMBL; AB037745; BAA92562.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 580 AA; 63365 MW; 5E63C19265EC8E5C CRC64;  
Query Match 2.0%; Score 21; DB 4; Length 580;  
Best Local Similarity 100.0%; Pred. No. 7.1e-13;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 943 LTCYFWKKNQKLEYKYSKLV M 963  
Db 502 LTCYFWKKNQKLEYKYSKLV M 522  
RESULT 4  
O82444  
ID O82444 PRELIMINARY; PRT; 552 AA.  
AC O82444;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Peroxisomal targeting sequence 1 receptor (Fragment).  
GN PEX5.  
OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99007315; PubMed=9789089;  
RA Kragler F., Lametschwandtner G., Christmann J., Hartig A.,  
RA Harada J.J.;  
RT "Identification and analysis of the plant peroxisomal targeting signal  
RT 1 receptor NtPEX5.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:13336-13341(1998).  
DR EMBL; AF056282; AAC69180.1; -.  
DR InterPro; IPR001440; TPR.  
DR Pfam; PF00515; TPR; 4.  
DR SMART; SM00028; TPR; 4.  
KW Receptor.  
FT NON\_TER 1  
SQ SEQUENCE 552 AA; 62293 MW; 7045FA177B0F51C6 CRC64;  
Query Match 0.9%; Score 9; DB 10; Length 552;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 RARGPVRGR 12  
Db 22 RARGPVRGR 30  
RESULT 5  
Q67608  
ID Q67608 PRELIMINARY; PRT; 55 AA.  
AC Q67608;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE Tomato golden mosaic virus subgenomic DNA derived from DNA B cccds -  
DE covalently closed circular double-stranded molecule.  
OS Tomato golden mosaic virus (TGMV).  
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.  
OX NCBI\_TaxID=10831;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87040767; PubMed=3022243;  
RA MacDowell S.W., Coutts R.H.A., Buck K.W.;  
RT "Molecular characterization of subgenomic single-stranded and double-  
RT stranded DNA forms isolated from plants infected with tomato golden  
RT mosaic virus.";  
RL Nucleic Acids Res. 14:7967-7984(1986).  
DR EMBL; X04485; CAA28171.1; -.



DR InterPro; IPR000211; Gemini\_BL.  
DR Pfam; PF00845; Gemini\_BL1; 1.  
SQ SEQUENCE 55 AA; 6281 MW; 41963570D739EC96 CRC64;  
  
Query Match 0.8%; Score 8; DB 12; Length 55;  
Best Local Similarity 100.0%; Pred. No. 5.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 177 NYIESNRD 184  
Db 13 NYIESNRD 20  
|||||  
  
RESULT 6  
Q9HR77 PRELIMINARY; PRT; 148 AA.  
AC Q9HR77;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Vng0825C.  
GN VNG0825C.  
OS Halobacterium sp. (strain NRC-1).  
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
OC Halobacteriaceae; Halobacterium.  
OX NCBI\_TaxID=64091;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20504483; PubMed=11016950;  
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,  
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welte R., Goo Y.A.,  
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,  
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;  
RT "Genome sequence of Halobacterium species NRC-1.";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
DR EMBL; AE005024; AAG19281.1; -.  
DR InterPro; IPR004435; MobB.  
DR Pfam; PF03205; MobB; 1.  
KW Complete proteome.  
SQ SEQUENCE 148 AA; 15073 MW; F242336129C362FE CRC64;  
  
Query Match 0.8%; Score 8; DB 17; Length 148;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 155 TVVGPSDS 162  
Db 5 TVVGPSDS 12  
|||||  
  
RESULT 7  
Q9DGZ5 PRELIMINARY; PRT; 154 AA.  
AC Q9DGZ5;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE X protein.  
GN X.  
OS Hepatitis B virus.  
OC Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.  
OX NCBI\_TaxID=10407;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=14118, AND 11141;  
RX PubMed=10950984;  
RA Hannoun C., Norder H., Lindh M.;  
RT "An aberrant genotype revealed in recombinant hepatitis B virus strains from Vietnam.";

RL J. Gen. Virol. 81:2267-2272(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=14118, AND 11141;  
RA Hannoun C., Norder H., Lindh M.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF241408; AAG17579.1; -.  
DR EMBL; AF241407; AAG17572.1; -.  
DR InterPro; IPR000236; TransactX.  
DR Pfam; PF00739; X; 1.  
SQ SEQUENCE 154 AA; 16516 MW; A13D1D623218640D CRC64;  
  
Query Match 0.8%; Score 8; DB 12; Length 154;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 48 GDLPSSSS 55  
Db 35 GDLPSSSS 42  
|||||  
  
RESULT 8  
Q9E6T5 PRELIMINARY; PRT; 154 AA.  
ID Q9E6T5  
AC Q9E6T5;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE X protein.  
GN X.  
OS Hepatitis B virus.  
OC Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.  
OX NCBI\_TaxID=10407;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=6871;  
RX MEDLINE=20409088; PubMed=10950984;  
RA Hannoun C., Norder H., Lindh M.;  
RT "An aberrant genotype revealed in recombinant hepatitis B virus strains from Vietnam.";  
RL J. Gen. Virol. 81:2267-2272(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=6871;-  
RA Hannoun C., Norder H., Lindh M.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF241409; AAG17586.1; -.  
DR InterPro; IPR000236; TransactX.  
DR Pfam; PF00739; X; 1.  
SQ SEQUENCE 154 AA; 16563 MW; E6B37B623218641B CRC64;  
  
Query Match 0.8%; Score 8; DB 12; Length 154;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 48 GDLPSSSS 55  
Db 35 GDLPSSSS 42  
|||||  
  
RESULT 9  
Q9ABT4 PRELIMINARY; PRT; 239 AA.  
ID Q9ABT4  
AC Q9ABT4;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein CC0136.  
GN CC0136.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
OC Caulobacter.  
OX NCBI\_TaxID=155892;



RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
RN [2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=BRISTOL N2;  
RC Tin-Wollam A.M.;  
RA "The sequence of C. elegans cosmid C43H8.";  
RT Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RN SEQUENCE FROM N.A.  
RP STRAIN=BRISTOL N2;  
RC Waterston R.;  
RA Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.  
RL EMBL; AF098499; AAC67398.1; -.  
DR InterPro; IPR001791; Laminin\_G..  
DR Pfam; PF00054; laminin\_G; 1.  
DR SMART; SM00282; LamG; 1.  
SQ SEQUENCE 281 AA; 31708 MW; 5253CC040277343C CRC64;  
  
Query Match 0.8%; Score 8; DB 5; Length 281;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 NITDFTVK 738  
Db 111 NITDFTVK 118

RESULT 13  
Q9PR14 PRELIMINARY; PRT; 286 AA.  
ID Q9PR14  
AC Q9PR14;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE ATP synthase gamma chain.  
GN ATPG OR UU130.  
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Ureaplasma.  
OX NCBI\_TaxID=134821;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SEROVAR 3;  
RX MEDLINE=20500219; PubMed=11048724;  
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,  
RA Cassell G.H.;  
RT "The complete sequence of the mucosal pathogen Ureaplasma  
RT urealyticum.";  
RL Nature 407:757-762(2000).  
DR EMBL; AE002114; AAF30536.1; -.  
DR InterPro; IPR000131; ATPase\_gamma.  
DR Pfam; PF00231; ATP-synt; 1.  
DR PRINTS; PR00126; ATPASEGAMMA.  
DR TIGRFAMs; TIGR01146; ATPsyn\_Flgamma; 1.  
KW Complete proteome.  
SQ SEQUENCE 286 AA; 32295 MW; C9AE278976AAE4F5 CRC64;

Query Match 0.8%; Score 8; DB 16; Length 286;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 EIVAGSDD 746  
Db 279 EIVAGSDD 286

RESULT 14

Q8V9II PRELIMINARY; PRT; 293 AA.  
ID Q8V9II  
AC Q8V9II;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Movement protein.  
GN MP.  
OS Squash yellow mottle virus.  
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.  
OX NCBI\_TaxID=91372;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ramirez P., Karkashian J., Zuniga C., Maxwell D.;  
RT "Association of Squash Yellow Mottle Virus with cucurbits and papaya  
RT in Costa Rica.";  
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AF440790; AAL33677.1; -.  
DR InterPro; IPR000211; Gemini\_BL.  
DR Pfam; PF00845; Gemini\_BLI; 1.  
SQ SEQUENCE 293 AA; 33078 MW; 8F99918F4FFE54AC CRC64;

Query Match 0.8%; Score 8; DB 12; Length 293;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 NYIESNRD 184  
Db 13 NYIESNRD 20

RESULT 15  
Q50198 PRELIMINARY; PRT; 379 AA.  
ID Q50198  
AC Q50198;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE L222-ORF9.  
OS Mycobacterium leprae.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1769;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97124199; PubMed=8969512;  
RA Fsihi H., De Rossi E., Salazar L., Cantoni R., Labo M., Riccardi G.,  
RA Takiff H.E., Eiglmeier K., Bergh S., Cole S.T.;  
RT "Gene arrangement and organization in a approximately 76 kb fragment  
RT encompassing the oric region of the chromosome of Mycobacterium  
RT leprae.";  
RL Microbiology 142:0-0(0).  
DR EMBL; L39923; AAB53129.1; -.  
DR InterPro; IPR004268; MVIN\_like.  
DR Pfam; PF03023; MVIN; 1.  
SQ SEQUENCE 379 AA; 40010 MW; FD23F9F20B4A74F5 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 379;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 TAVLLVAL 943  
Db 243 TAVLLVAL 250

Search completed: May 12, 2003, 13:40:54  
Job time : 35.5452 secs





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 07:48:38 ; Search time 49.5437 Seconds  
(without alignments)  
2762.174 Million cell updates/sec

Title: US-10-073-333A-4  
Perfect score: 5681  
Sequence: 1 MLFRARGPVRRGWGRPAEA.....KEKEDHESVQLKTSRSPNI 1027

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5681	100.0	1027	22 AAB70256	TR16-long receptor
2	5307	93.4	963	22 AAB70255	TR16-short recepto
3	3050.5	53.7	1013	21 AAB26179	Human CASB619 prot
4	3040.5	53.5	1013	22 AAB3845	Amino acid sequenc
5	3025	53.2	1013	22 AAU12190	Human PRO4985 poly
6	2982.5	52.5	1001	22 AAB35333	Human TR13 recepto
7	2712.5	47.7	911	22 AAB83850	Amino acid sequenc
8	2702.5	47.6	870	22 AAB83851	Amino acid sequenc
9	2526	44.5	464	22 AAB48377	Human SEC10 protei
10	2222	39.1	411	22 AAB48372	Human SEC5 protein

11	2167	38.1	750	22 AAB35328	Human TR13 recepto
12	1487.5	26.2	495	20 AAY59972	Human endometrium
13	1208.5	21.3	383	22 AAB83853	Amino acid sequenc
14	883	15.5	372	22 AAB85768	Human seven-transm
15	705.5	12.4	209	22 AAB83852	Amino acid sequenc
16	452	8.0	81	22 ABB39918	Peptide #7424 enco
17	452	8.0	81	22 ABB24471	Protein #6470 enco
18	452	8.0	81	22 AAM60663	Human brain expres
19	452	8.0	81	22 AAM73335	Human bone marrow
20	452	8.0	81	22 AAM33535	Peptide #7572 enco
21	452	8.0	81	23 ABG43186	Human peptide enco
22	444.5	7.8	208	21 AAB53442	Human colon cancer
23	442	7.8	74	22 AAB70281	Peptide #25. Unid
24	395	7.0	78	22 AAB70285	Peptide #29. Unid
25	370.5	6.5	147	22 AAB83849	Peptide fragment o
26	353	6.2	60	22 ABB38686	Peptide #6192 enco
27	353	6.2	60	22 ABB23759	Protein #5758 enco
28	353	6.2	60	22 AAM59318	Human brain expres
29	353	6.2	60	22 AAM71867	Human bone marrow
30	353	6.2	60	22 AAM32149	Peptide #6186 enco
31	353	6.2	60	23 ABG41680	Human peptide enco
32	353	6.2	64	22 AAM72925	Human bone marrow
33	353	6.2	64	23 ABG42760	Human peptide enco
34	353	6.2	105	21 AAB26180	Human CASB619 prot
35	350	6.2	150	20 AAY12274	Human 5' EST secre
36	322	5.7	71	22 AAU21345	Human novel foetal
37	255	4.5	50	22 ABB39681	Peptide #7187 enco
38	255	4.5	50	22 ABB24346	Protein #6345 enco
39	255	4.5	50	22 AAM60397	Human brain expres
40	255	4.5	50	22 AAM73033	Human bone marrow
41	255	4.5	50	22 AAM19811	Peptide #6245 enco
42	255	4.5	50	22 AAM33257	Peptide #7294 enco
43	255	4.5	50	23 ABG42877	Human peptide enco
44	205	3.6	1605	21 AAB19805	Mouse laminin 2 ga
45	205	3.6	1605	21 AAB48454	Mouse laminin 8 po

ALIGNMENTS

RESULT 1  
AAB70256  
ID AAB70256 standard; protein; 1027 AA.  
XX  
AC AAB70256;  
XX  
DT 10-MAY-2001 (first entry)  
XX  
DE TR16-long receptor protein.  
XX  
KW TR16 recptor; tumour necrosis factor receptor superfamily;  
KW apoptosis; inflammatory; cancer; immune; neurodegenerative.  
XX  
OS Unidentified.  
XX  
PN WO200112671-A1.  
XX  
PD 22-FEB-2001.  
XX  
PF 10-AUG-2000; 2000WO-US21885.  
XX  
PR 12-AUG-1999; 99US-0148348.  
PR 13-AUG-1999; 99US-0148683.  
PR 13-AUG-1999; 99US-0148870.  
PR 16-AUG-1999; 99US-0148758.  
PR 17-AUG-1999; 99US-0149181.  
PR 18-AUG-1999; 99US-0149453.  
PR 19-AUG-1999; 99US-0149498.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Young PE, Baker KP;  
XX

DR	WPI; 2001-138754/14.									
XX	New nucleic acid molecule encoding a TR16 tumor necrosis factor									
PT	receptor polypeptide, useful for the diagnosis and treatment of cancer,									
PT	autoimmune disorders and cardiovascular diseases -									
XX										
PS	Disclosure; Fig 4; 286pp; English.									
XX										
CC	The present invention relates to a TR16 receptor (tumour necrosis									
CC	factor receptor superfamily). The invention is useful treating									
CC	diseases and disorders associated with the inhibited or increased									
CC	apoptosis. In particular inflammatory diseases, cancers, immune and									
CC	neurodegenerative disorders may be treated.									
XX										
XX										
QY	Sequence	1027 AA;								
	Query Match	100.0%;	Score	5681;	DB	22;	Length	1027;		
	Best Local Similarity	100.0%;	Pred.	NO.	0;					
	Matches 1027;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;	
QY	1	MLFRARGPVGRGWGRPAEAPRRGRSPWPSPAWICWALAGCAAWAGDLPSSSSRPLPP	60							
Db	1	MLFRARGPVGRGWGRPAEAPRRGRSPWPSPAWICWALAGCAAWAGDLPSSSSRPLPP	60							
QY	61	COEKDYHFEYTECDSSGSRWRVAIPNSAVDC\$GLPDVVRGKECTFSCASGEYLEMKNQVC	120							
Db	61	COEKDYHFEYTECDSSGSRWRVAIPNSAVDC\$GLPDVVRGKECTFSCASGEYLEMKNQVC	120							
QY	121	SKCGEGTSLGSGIKFDEWDELPAFESNIATEFMDTVVGPDSRDPGCNNSWIIPRGNYIE	180							
Db	121	SKCGEGTSLGSGIKFDEWDELPAFESNIATEFMDTVVGPDSRDPGCNNSWIIPRGNYIE	180							
QY	181	SNRDDCTVSLIYAVHLKKGYSYVFEFYQYVDNNIFFEYQNDQCEQEMDTTDDKWKVCLTDN	240							
Db	181	SNRDDCTVSLIYAVHLKKGYSYVFEFYQYVDNNIFFEYQNDQCEQEMDTTDDKWKVCLTDN	240							
QY	241	GEWGSHSVMLKSGTNILYWRTTGILMGSKAVKPVLVKNITIEGVAYTSECFFCPKPGTFSN	300							
Db	241	GEWGSHSVMLKSGTNILYWRTTGILMGSKAVKPVLVKNITIEGVAYTSECFFCPKPGTFSN	300							
QY	301	KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG	360							
Db	301	KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG	360							
QY	361	KTQIMYKWIPIKICREDLTDAILRPPSGEKKDCPPCPNPGFYNNSSSCHPCPPGTFSDGT	420							
Db	361	KTQIMYKWIPIKICREDLTDAILRPPSGEKKDCPPCPNPGFYNNSSSCHPCPPGTFSDGT	420							
QY	421	KECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS	480							
Db	421	KECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS	480							
QY	481	DNDYLILNLHIPGFKPPTSMGTATGSELGRITFVFETLCSADCVLVFMVDINRKSTNVVE	540							
Db	481	DNDYLILNLHIPGFKPPTSMGTATGSELGRITFVFETLCSADCVLVFMVDINRKSTNVVE	540							
QY	541	SWGGTKKQAYTHIIFKNATFTTWAFTNQGDNRFRINDMVKIYSITATNAVDGVAS	600							
Db	541	SWGGTKKQAYTHIIFKNATFTTWAFTNQGDNRFRINDMVKIYSITATNAVDGVAS	600							
QY	601	SCRACALGSEQSSCVPCPPGHYIEKETNOCKECPDPTYLSIHQVYGKEACIPCGPGSK	660							
Db	601	SCRACALGSEQSSCVPCPPGHYIEKETNOCKECPDPTYLSIHQVYGKEACIPCGPGSK	660							
QY	661	NNQDHSVCSYDCFFYHEKENQILHYDFSLSVSGSLMNGPSETSCKTKYFHHFNISLCGH	720							
Db	661	NNQDHSVCSYDCFFYHEKENQILHYDFSLSVSGSLMNGPSETSCKTKYFHHFNISLCGH	720							
QY	721	EGKKNALCTNNITDFTVKEIVAGSDDYTNLVGAFVCSQSTIIPSESKGFRAALSSQSIIILA	780							
Db	721	EGKKNALCTNNITDFTVKEIVAGSDDYTNLVGAFVCSQSTIIPSESKGFRAALSSQSIIILA	780							
QY	781	DTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSSTATTSINGRSTAVKMRCNPT	840							

Db	781	DTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSATTSCINGRSTAVKMRCNPT	840	
QY	841	KSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEHDHFEIEGACKRGFQETLYVWN	900	
Db	841	KSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEHDHFEIEGACKRGFQETLYVWN	900	
QY	901	EPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAGTAVLLVALTCYFWKKKNOKLEYKYSK	960	
Db	901	EPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAGTAVLLVALTCYFWKKKNOKLEYKYSK	960	
QY	961	LVMTTNSKECELPAAADSCAIMEGEDNEEEVVYSNKQSLGKLSLATKEKEDHFESVQLK	1020	
Db	961	LVMTTNSKECELPAAADSCAIMEGEDNEEEVVYSNKQSLGKLSLATKEKEDHFESVQLK	1020	
QY	1021	TSRSPNI 1027		
Db	1021	TSRSPNI 1027		
RESULT 2				
AAB70255				
ID	AAB70255 standard; protein; 963 AA.			
XX	AAB70255;			
AC	AAB70255;			
XX	10-MAY-2001 (first entry)			
DT	TR16-short receptor protein.			
XX	TR16 receptor; tumour necrosis factor receptor superfamily;			
DE	apoptosis; inflammatory; cancer; immune; neurodegenerative.			
XX	Unidentified.			
OS	WO200112671-A1.			
XX	22-FEB-2001.			
PN	10-AUG-2000; 2000WO-US21885.			
XX	12-AUG-1999; 99US-0148348.			
PR	13-AUG-1999; 99US-0148683.			
PR	13-AUG-1999; 99US-0148870.			
PR	16-AUG-1999; 99US-0148758.			
PR	17-AUG-1999; 99US-0149181.			
PR	18-AUG-1999; 99US-0149453.			
PR	19-AUG-1999; 99US-0149498.			
XX	(HUMA-) HUMAN GENOME SCI INC.			
PA	Ruben SM, Young PE, Baker KP;			
XX	WPI; 2001-138754/14.			
DR	New nucleic acid molecule encoding a TR16 tumor necrosis factor			
XX	receptor polypeptide, useful for the diagnosis and treatment of cancer			
PT	autoimmune disorders and cardiovascular diseases -			
PT	Claim 1; Fig 1; 286pp; English.			
XX	The present invention relates to a TR16 receptor (tumour necrosis			
CC	factor receptor superfamily). The invention is useful treating			
CC	diseases and disorders associated with the inhibited or increased			
CC	apoptosis. In particular inflammatory diseases, cancers, immune and			
CC	neurodegenerative disorders may be treated.			
XX	Sequence 963 AA;			
SQ	Query Match 93.4%; Score 5307; DB 22; Length 963;			
Best Local Similarity 100.0%; Pred. No. 0;				
Matches 953; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				

```
QY 1 MLFRARGPVRGRGWGRPAEAPRRGRSPWPSPAWICCWALAGCQAAGDLPSSSSRPLPP 60
  |||
Db 1 MLFRARGPVRGRGWGRPAEAPRRGRSPWPSPAWICCWALAGCQAAGDLPSSSSRPLPP 60

QY 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFCASGEYLEMKNQVC 120
  |||
Db 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFCASGEYLEMKNQVC 120

QY 121 SKCGEGTYSLGSGIKFDEWDELPAAGFSNIATFMDTVVGPDSRDPGCNNSWIPRGNIE 180
  |||
Db 121 SKCGEGTYSLGSGIKFDEWDELPAAGFSNIATFMDTVVGPDSRDPGCNNSWIPRGNIE 180

QY 181 SNRDDCTVSLIYAVHLKSGYVFFEQYVDNNIFFEFFIIONDQCQEMDTTIDKWKLTDN 240
  |||
Db 181 SNRDDCTVSLIYAVHLKSGYVFFEQYVDNNIFFEFFIIONDQCQEMDTTIDKWKLTDN 240

QY 241 GEWGSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFFPKPGTFSN 300
  |||
Db 241 GEWGSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFFPKPGTFSN 300

QY 301 KPGSFNCQVCPNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360
  |||
Db 301 KPGSFNCQVCPNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360

QY 361 KTOIMYKWIPIKICREDLTDAIRLPSGEKKDCPPCNPFGFYNNSSSCHPCPPGTFSDGT 420
  |||
Db 361 KTOIMYKWIPIKICREDLTDAIRLPSGEKKDCPPCNPFGFYNNSSSCHPCPPGTFSDGT 420

QY 421 KECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480
  |||
Db 421 KECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480

QY 481 DNDYLILNLHIPGFKPPTSMTGATGSELGRITFVFETLCSADCVLFMVDINRKSTNVVE 540
  |||
Db 481 DNDYLILNLHIPGFKPPTSMTGATGSELGRITFVFETLCSADCVLFMVDINRKSTNVVE 540

QY 541 SWGGTKEKQAYTHIIFKNATFTFWAFQRTNQGDNRFRINDVMKIYSITATNAVGVAS 600
  |||
Db 541 SWGGTKEKQAYTHIIFKNATFTFWAFQRTNQGDNRFRINDVMKIYSITATNAVGVAS 600

QY 601 SCRACALGSEQSGSSCVPCPPGPHYIEKETNQCKECPDPTYLSIHQVYGKEACIPCGPGSK 660
  |||
Db 601 SCRACALGSEQSGSSCVPCPPGPHYIEKETNQCKECPDPTYLSIHQVYGKEACIPCGPGSK 660

QY 661 NNQDHSVCSYDCFFYHEKENQILHYDFSNLSSVGLMNGPFTSKGTKYFHFNNISLCGH 720
  |||
Db 661 NNQDHSVCSYDCFFYHEKENQILHYDFSNLSSVGLMNGPFTSKGTKYFHFNNISLCGH 720

QY 721 EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFVCQSTIIPSESKGFRAALSSQSIILA 780
  |||
Db 721 EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFVCQSTIIPSESKGFRAALSSQSIILA 780

QY 781 DTFIGVTVETTLKNINIKEDMPVPPTSQIPDVHFFYKSTATTSCINGRSTAVKMRGNPT 840
  |||
Db 781 DTFIGVTVETTLKNINIKEDMPVPPTSQIPDVHFFYKSTATTSCINGRSTAVKMRGNPT 840

QY 841 KSGAGVISVPSKCPAGTCDCGCTFYFLWESAECPLCTEHDFFEIEGACKRGFQETLYVWN 900
  |||
Db 841 KSGAGVISVPSKCPAGTCDCGCTFYFLWESAECPLCTEHDFFEIEGACKRGFQETLYVWN 900

QY 901 EPKWCIGISLPEKKLATCETVDFWLKVGAGVGAFATVLLVALTCYFWKKNQK 953
  |||
Db 901 EPKWCIGISLPEKKLATCETVDFWLKVGAGVGAFATVLLVALTCYFWKKNQK 953
```

RESULT 3  
AAB26179  
ID AAB26179 standard; Protein; 1013 AA.  
XX

AC AAB26179;  
XX  
DT 12-FEB-2001 (first entry)  
XX

```
DE Human CASB619 protein #1.
XX
KW Human; CASB619; cancer; autoimmune disease; immunogen; vaccine;
  epitope.
XX
OS Homo sapiens.
XX
PN WO200058460-A2.
XX
PD 05-OCT-2000.
XX
PF 20-MAR-2000; 2000WO-EP02478.
XX
PR 26-MAR-1999; 99GB-0007113.
PR 25-SEP-1999; 99GB-0022858.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PI Brück CEM, Cassart J, Coche T, Vinals De Bassols YC;
XX
DR WPI; 2000-664923/64.
DR N-PSDB; AAA95442.
XX
PT Novel CASB619 polypeptides useful for diagnosing, and as vaccines for
  prophylactic and therapeutic treatment of, cancers, particularly
  ovarian and colon carcinoma, and autoimmune diseases
XX
PS Claim 4; Page 54-56; 68pp; English.
XX
CC The present sequence comprises the human CASB619 protein sequence. This
  protein is thought to be specifically or over-expressed in tumour cells,
  and so can be used as a target for antigen-specific immune responses
  which can cause destruction of the tumour cell. In addition, the protein
  and gene can be used in cancer diagnosis, in the treatment of autoimmune
  diseases and in vaccines against cancer and autoimmune disease. The
  invention provides a number of epitopes derived from the protein which
  can be used as immunogens.
XX
SQ Sequence 1013 AA;
```

Query Match 53.7%; Score 3050.5; DB 21; Length 1013;  
Best Local Similarity 54.0%; Pred. No. 6.9e-231;  
Matches 543; Conservative 171; Mismatches 271; Indels 21; Gaps 11;

```
QY 23 RGRSPWPSPAWICCWALAGCQAAG--DLPSSSSRPLPCQEKDYHFEYTECDSSGSR 79
  |||
Db 15 RGRTERRIPR---LWRL---LLWAGTAFQVTQGTGPELHACKESEYHYEYTACDSTGSR 67

QY 80 WRVAIPNSAVDCSGLPDPVRGKECTFCASGEYLEMKNQVCSKCGEGTYSLGSGIKFDEW 139
  |||
Db 68 WRVAVPTPGLCTSLPDPVKGTCECSFCNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEW 127

QY 140 DELPAGFSNIATFMDTVVGPDSRDPGCNNSWIPRGNIESNRDDCTVSLIYAVHLKKS 199
  |||
Db 128 DELPHGFASLSANMELDDSAES-TGNCTSSKWVPRGDYIASNTDECTATLMYAVNLKQS 186

QY 200 GYVFFEYQYVDNNIFFEFFIIONDQCQEMDTTIDKWKVLTNGEWGSHSVMLKSGTNILYW 259
  |||
Db 187 GTVNFYIYYPDSSIIIEFFVQNDQCOP-NADDSRWMTTEKG-WEFHSVELNRGNVLYW 244

QY 260 RTTGILMGSKAVKPVLVKNITIEGVAYTSECFFPKPGTFSNKPFGSFCNQCVRNTYSEKG 319
  |||
Db 245 RTTAFSVWTKVPKPVLVNRNIATGVAYTSECFFPKPGTYADKQGSFCKLCPANSYSNKG 304

QY 320 AKECIRCKDDSQFS--GSSECTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIPIKICRED 377
  |||
Db 305 ETSCHQC-DPDKYSEKSGSSCNVRPACTDKDYFTHFACDANGETQLMYKWAKPKICSED 363

QY 378 LTDAIRLPPSGEKKDCPPCNPFGFYNNSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFE 437
  |||
Db 364 LEGAVKLPASGVKTHCPCNPGFFKTNNSTCQPCPCYGSYNGS-DCTRCPAGTEPAVGFE 422

QY 438 YKWNVNLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSNDYLIILNLHIPGFKPP 497
```



Db 423 YKWNLTPTNMETTVLSGINFYKGMTGWEVAGDHIYTAAGASDNDFMILTLVPGFRPP 482  
Qy 498 TS-MTGATGSELGRITTFVETLCSADCVLVYFMDINRSTNVVSWGGTKEKQAYTHIF 556  
Db 483 QSVMDATENKEVARITTFVETLCSVNCLELYFVMGVNSRTPVETWKGSKGQSYTHIE 542  
Qy 557 KNATFTFTWAFQRTNQGDNRFRINDMVKIYSITATNAVGVASSCRACALGSEQSGSSC 616  
Db 543 ENTTSFTWAFQRTTFHEASRKYTNDVAKIYSINVTNMNGVASYCRPCALEASDVGSSC 602  
Qy 617 VPCPPGHVIEKETNOCKECPDPTYSIHQVYGEACIPCGPGSKNNQDHSVCYSDCFYH 676  
Db 603 TSCPAGYIDRDSGTCHSCPPNTILKAHQPYGVQACVPCPGTKNKKIHSCLYNDCTFSR 662  
Qy 677 EKENQILHYDFSNLSSVGLMNGSPFTSKGTYFFHFNISLCGHEGKKMALCTNNITDFT 736  
Db 663 NTPTRTFYNFESALANTVTLAGSPFTSKGLKYFHHFTLSLCGNQGRKMSVCTDNVTDLR 722  
Qy 737 VKEIVAGSDDYTNLVGAFVCQSTIIPSESKGFRAALSSQSILADTFIGVTVETTLKNIN 796  
Db 723 IPE--GSGFSKSITAYVCQAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGIT 779  
Qy 797 IKEDMPFPVPTSQIPDVHFFYKSTATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAG 856  
Db 780 SPAELFHLESGLIPDVIFFYRSNDVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCSDG 839  
Qy 857 TCDGCTFYFLWESAEACPLCTEHDHFEIAGACKRGFQETLYVMNEPKWCIKGISLPEKKL 916  
Db 840 TCDGCNFHLEWESAAACPLCSVADYHAIVSSCVAGIQKTTYVWREPKLCSGGISLPEQRV 899  
Qy 917 ATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKNQKLEYKYSKLVMTNSKECELPAAD 976  
Db 900 TICKTIDFWLKVGISAGTCTAILLTVLTCYFWKKNQKLEYKYSKLVMTNLKDCDLPAAD 959  
Qy 977 SCAIMEGEDNEEVVYSNQSLLGLKLSLATKEKEDHFESVOLKTS 1022  
Db 960 SCAIMEGEDVEDDLIFTSKSLFGKIKSFTSKRTPDGFDSVPLKTS 1005  
RESULT 4  
AAB83845  
ID AAB83845 standard; Protein; 1013 AA.  
XX  
AC AAB83845;  
XX  
XX  
DT 23-JUL-2001 (first entry)  
XX  
DE Amino acid sequence of a human protein expressed in tumour cells.  
XX  
DE Tumour cell; immunological disease; autoimmune disease; cancer;  
KW Infection.  
KW  
XX  
OS Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..41  
FT /note= "signal peptide"  
FT Domain 42..911  
FT /note= "extracellular domain"  
FT Domain 912..930  
FT /note= "transmembrane domain"  
FT Domain 931..1013  
FT /note= "transmembrane domain"  
XX  
XX WO200131003-A1.  
XX  
XX  
PD 03-MAY-2001.  
XX  
XX 30-OCT-2000; 2000WO-FR03032.  
XX  
XX 29-OCT-1999; 99FR-0013629.  
XX

PA (FABR ) FABRE MEDICAMENT SA PIERRE.  
XX  
PI Delneste Y, Magistrelli G, Jeannin P, Bonnefoy J;  
XX  
DR WPI; 2001-328651/34.  
DR N-PSDB; AAF89765.  
XX  
PT New nucleic acid, expressed in tumours and lymphoid tissue is useful for  
PT identifying agents for treating tumours and autoimmune disease -  
XX  
PS Claim 9; Page 48-51; 85pp; French.  
XX  
CC The present sequence represents a human protein expressed in tumour  
CC cells. The polynucleotide is useful for screening cDNA/genomic DNA banks  
CC and for cloning isolated DNA; identifying mutant forms of the gene that  
CC encodes a human protein, where the mutations are associated with  
CC abnormal gene expression, or promoters and regulators of the gene,  
CC particularly for diagnosis; for recombinant expression of the derived  
CC protein; as probes and primers for detection and amplification; and  
CC as antisense therapeutics. The tumour expressed protein is useful for  
CC raising specific antibodies and to screen agents that modulate its  
CC activity, bind to it or interact with it. These agents are potentially  
CC useful for treatment or prevention of diseases associated with abnormal  
CC expression/activity of the protein, particularly immunological diseases  
CC (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic  
CC infections.  
XX  
SQ Sequence 1013 AA;  
Query Match 53.5%; Score 3040.5; DB 22; Length 1013;  
Best Local Similarity 53.9%; Pred. No. 4.2e-230;  
Matches 542; Conservative 170; Mismatches 273; Indels 21; Gaps 11;  
Qy 23 RGRSPWPSPAWICCWALAGCQAAWAG--DLPSSSRPLPPCQEKDYHFEYTECDSSGSR 79  
Db 15 RGRTERIPR---LWRL---LLWAGTAFQVQTGTGPELHACKESEYHYEYACDSYSGSR 67  
Qy 80 WRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVCKGEGTYSLSGSGIKFDEW 139  
Db 68 WRVAVPHTPGLCTSLPDPVKGTECSFSCNAGEFLDMKDQCKPKCAEGRYSLSGTGIRFDEW 127  
Qy 140 DELPAGESNIATFMDTVVGPSPDSRDPDCNNSWIPRGNYIESNRDDCTVSLIYAVHLKKS 199  
Db 128 DELPHGFASLSANMELDDSAAES-TGNCSTSSKVVPRGDYIASNTDECTATLMYAVNLKQS 186  
Qy 200 GYVFEFEYQYVDNNIEFFEQNDQCEMDTTTDKWKVLTDNGEWSHVMKSGTNILYW 259  
Db 187 GTVNFEEYYPDSSIIFFVQNDQCP-NADDSRWKMTTEKG-WEFHSVELNRGNVLYW 244  
Qy 260 RTTGILMCSKAVKPVLVKNITIEGVAYTSECFCKPGTFSNKPQSFNCQVCPRNTYSEKG 319  
Db 245 RTTAFSVWTKVPKPVLRNIAITGVAYTSECFCKPGTGYADKQSSFCCKLPANSYSNKG 304  
Qy 320 AKECIRCKDDSQFS--GSSECTERPPCTTKDYFIHTPCDEEGKTOIMYKWIETPKICRED 377  
Db 305 ETSCHQC-DPDKYSEKSGSSCNVRPACTDKDYFYTHTACDANGEIQLMYKWKAPKICSED 363  
Qy 378 LTDAIRLPPSGEKKDCPPCNPGFYNNGSSSCHIPCPTGTFSDGTKECRPCPAGTEPALGFE 437  
Db 364 LEGAVKLPAASGVKTHPCPNPGFEKTNNSTCQPCPYGPNYSNGS-DCTRCPCPAGTEPALGFE 422  
Qy 438 YKWNVLPGNMKTSFCFNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPP 497  
Db 423 YKWNLTPTNMETTVLSGINFEYKGMTGWEVAGDHIYTAAGASDNDFMILTLVPGFRPP 482  
Qy 498 TS-MTGATGSELGRITTFVETLCSADCVLVYFMDINRSTNVVSWGGTKEKQAYTHIF 556  
Db 483 QSVMDATENKEVARITTFVETLCSVNCLELYFVMGVNSRTPVETWKGSKGQSYTHIE 542  
Qy 557 KNATFTFTWAFQRTNQGDNRFRINDMVKIYSITATNAVGVASSCRACALGSEQSGSSC 616  
Db 543 ENTTSFTWAFQRTTFHEASRKYTNDVAKIYSINVTNMNGVASYCRPCALEASDVGSSC 602

QY 617 VPCPPGHYIEKETNOCKECPDPTLYLSIHQVYGEACIPCGPGSKNNQDHSVCYSDCFFYH 676  
|||:||||:| |||:| | |||:||||:||||: |||:| | |||:| |  
Db 603 TSCPAGYIIDRDSGTCHSCPNTILKAHQPYGVQACVPCGPGTKNNKIHSCLYNDCTFSR 662  
QY 677 EKENQILHYDEFNLSSVGLMNGPSFTSKGTKYHFFFNISLCGHEGKKMALCTNNITDFT 736  
: |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| |  
Db 663 NTPTRTFNYNFSALANTVTLAGGSPFTSKGLYFHHTLSLCCGNQGRKMSVCTDNVTDLR 722  
QY 737 VKEIVAGSDDYTNLVGAFVCQSTIIPSESKGFRALSSQSIIADTFIGVTVETTLKNIN 796  
: | | : : |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| |  
Db 723 IPE---GESGFSKSIAYVCQAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGIT 779  
QY 797 IKEDMFPVPTSQIPDVHFFYKSSATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAG 856  
: | : |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| |  
Db 780 SPAELFHLESGLIPDVIFFYRSNDVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCSDG 839  
QY 857 TCDGCTFYFLWESAECPLCTEHDFHEIEGACKRGFQETLYVWNEPKWCIKGISLPEKKL 916  
|||||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| |  
Db 840 TCDGCNHFHFLWESAACPLCSVADYRAIVSSCVAGIQKTTYVWREPKLCSGGISLPEQRV 899  
QY 917 ATCETVDFWLKVGAGYGAVTAALLVALTCYFWKKNQKLEYKYSKLVMTTNSKECELPAAD 976  
|:|:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| |  
Db 900 TICKTIDFWLKVGISAGCTAILLTVLTCTCYFWKKNQKLEYKYSKLVMTNATLKDCDLPAAD 959  
QY 977 SCAIMEGEDNEEEVYSNQSLLGKLKSLATKEKEDHFESVLKTS 1022  
|||||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| |  
Db 960 SCAIMEGEDVEDDLIFTSKSLFKGKIKSFTSKRTPDGFDSVPLKTS 1005  
RESULT 5  
AAU12190  
ID AAU12190 standard; Protein; 1013 AA.  
XX AC AAU12190;  
XX DT 24-OCT-2001 (first entry)  
DE Human PRO4985 polypeptide sequence.  
XX KW Human secretory and transmembrane; PRO; mammalian; cancer; lung;  
KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;  
KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;  
KW adipocyte; A-peptide; factor VIIA; gene therapy.  
XX OS Homo sapiens.  
XX PN WO200140466-A2.  
XX PD 07-JUN-2001.  
XX PF 01-DEC-2000; 2000WO-US32678.  
XX PR 01-DEC-1999; 99WO-US28301.  
PR 01-DEC-1999; 99WO-US28634.  
PR 02-DEC-1999; 99WO-US28551.  
PR 02-DEC-1999; 99WO-US28564.  
PR 02-DEC-1999; 99WO-US28565.  
PR 09-DEC-1999; 99US-0170262.  
PR 16-DEC-1999; 99WO-US30095.  
PR 20-DEC-1999; 99WO-US30911.  
PR 20-DEC-1999; 99WO-US30999.  
PR 30-DEC-1999; 99WO-US31243.  
PR 06-JAN-2000; 2000WO-US00277.  
PR 06-JAN-2000; 2000WO-US00376.  
PR 11-FEB-2000; 2000WO-US03565.  
PR 18-FEB-2000; 2000WO-US04341.  
PR 18-FEB-2000; 2000WO-US04342.  
PR 22-FEB-2000; 2000WO-US04414.  
PR 24-FEB-2000; 2000WO-US04914.  
PR 24-FEB-2000; 2000WO-US05004.  
PR 01-MAR-2000; 2000WO-US05601.  
PR 20-MAR-2000; 2000WO-US07377.  
PR 21-MAR-2000; 2000WO-US07532.

PR 30-MAR-2000; 2000WO-US08439.  
PR 17-MAY-2000; 2000WO-US13705.  
PR 22-MAY-2000; 2000WO-US14042.  
PR 30-MAY-2000; 2000WO-US14941.  
PR 02-JUN-2000; 2000WO-US15264.  
PR 10-NOV-2000; 2000WO-US30873.  
XX (GETH ) GENENTECH INC.  
PA Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;  
XX Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;  
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;  
XX WPI; 2001-408281/43.  
DR N-PSDB; AAS21262.  
XX Isolated, secretory and transmembrane PRO polypeptide used to detect  
PT other PRO polypeptides, link bioactive molecules to cells expressing  
PT PRO polypeptides, and detect the presence of mammalian tumours e.g.  
PT lung, breast, prostate, cervical  
XX Claim 12; Fig 38; 813pp; English.  
PS AAU12172-AAU12446 represent novel human secretory and transmembrane  
XX PRO polypeptides. The PRO polypeptides are useful to detect other  
CC PRO polypeptides, to link bioactive molecules to cells expressing  
CC PRO polypeptides, to modulate biological activities of cells expressing  
CC PRO polypeptides, and to detect the presence of mammalian lung, colon,  
CC breast, prostate, rectal, cervical or liver tumours by comparing PRO  
CC polypeptide expression in a cell sample to that in a control sample.  
CC Some of the 275 sequences are also useful to stimulate the release of  
CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the  
CC proliferation or differentiation of chondrocytes, the proliferation or  
CC gene expression in pericyte cells, the release of proteoglycans from  
CC cartilage, the proliferation of inner ear utricular supporting cells or  
CC of T-lymphocytes, the release of a cytokine from peripheral blood  
CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of  
CC the PRO polypeptides may modulate glucose or free fatty acid uptake by  
CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide  
CC to factor VIIA. The PRO polypeptides can be used in assays to identify  
CC molecules involved in binding interactions. The polynucleotides encoding  
CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,  
CC transgenic or knock out animals and can be used in gene therapy.  
XX SQ Sequence 1013 AA;  
Query Match 53.2%; Score 3025; DB 22; Length 1013;  
Best Local Similarity 54.4%; Pred. NO. 7e-229;  
Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;  
QY 46 WAG---DLPSSSRPLPPCQEKDYHFYEYTECDSSGSRWRVAIPNSAVDCSGLPDBVRGKE 102  
||| : | | : |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| |  
Db 31 WAGTAFQVTQGTGPELHACKESEYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKGTGTE 90  
QY 103 CTFSCASGEYLEMKNQVCSKCGEGTYSLSGSGIKFEDWDELPAFNSNIATFMDTVVGPSDS 162  
|:|:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| |  
Db 91 CSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIREDEWDELPHGFASLSANMELDDSAAES 150  
QY 163 RPDGCNSSWIPRGNVIESNRDDCTVSLIYAVHLKSGYVFFEYQYVDNNIFFEYFIQND 222  
| : | | : |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| |  
Db 151 -TGNCTSSKWVPRGDYIASNTDECTATLMYAVNLKSGTVNFYYPDSSIIEFFVQND 209  
QY 223 QCQEMDTTDDKWKVLTNDNGEWSHVMKSGTNILYWRRTTGILMGSKAVKPVLVKNITIE 282  
||| : : | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 210 QCQP-NADDSRWMTTEKG-WEFHVELNRGNVLYWRRTAFSVWTKVPKPVLVNRNIAIT 267  
QY 283 GVAYTSECFPCPKPGFNSNKPFSNQCVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTE 340  
|||||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| | |||:| |  
Db 268 GVAYTSECFPCPKPGTYADKQSSFCCLKCPANSYSNKGETSCHQC-DPDKYSEKSSSCNV 326  
QY 341 RPPCTTKDYFIHTPCDEEGKTQIMYKWIEPKICREDLTDAILRPPSGEKKDCPPCNPGE 400  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 327 RPACTDKDYFYHTACDANGETQLMYKWAKPKICSEDELEGAVKLPASGVKTHCPCPNPGE 386





QY	737	VKEIVAGSDDYTNLVGAFVCQSTIIIPSESKGFRAALSSQSIILADTFIGTVTETTLKNIN	796.
Db	723	IPE---GEGFSKSITAYVCQAVIIPPEVTGYKAGVSSQPVSLADRLIGVTMTLDGIT	779
QY	797	IKEDMFPVPTSIQIPDVHFFYKSSSTATTSCTINGRSTAVKMRNPTKSGAGVISVPSKCPAG	856
Db	780	SPAELFHLESLGIPDVIFFYRSNDVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCSDG	839
QY	857	TCDGCTFYFLWESAEACPLCTEHDFHEIEGACKRGFQETLYVWNEPKWCIGISLPEKL	916
Db	840	TCDGCNHFHFLWESAAACPLCSVADYHAIVSSCVAGIQKTTYVWREPKLCSGGISLPEQRV	899
QY	917	ATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKNQKLEYKYSKLVMTTNSKECLPAAD	976
Db	900	TICKTIDFWLKVGISAGTCTAILLTVLTCYFWKKNQKLEYKYSKLVMTTNSKECLPAAD	959
QY	977	SCAIMEGEDNEEVVYSNKQSLGLKLSL	1005
Db	960	SCAIMEGEDVEDDLIFTSKNHSLGRSNHL	988
RESULT 7			
AAB83850			
ID	AAB83850 standard; Protein; 911 AA.		
XX	AAB83850;		
XX			
DT	23-JUL-2001 (first entry)		
XX			
DE	Amino acid sequence of a human protein expressed in tumour cells.		
XX			
KW	Tumour cell; immunological disease; autoimmune disease; cancer;		
KW	infection.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200131003-A1.		
XX			
PD	03-MAY-2001.		
XX			
PF	30-OCT-2000; 2000WO-FR03032.		
XX			
PR	29-OCT-1999; 99FR-0013629.		
XX			
PA	(FABR ) FABRE MEDICAMENT SA PIERRE.		
XX			
PI	Delneste Y, Magistrelli G, Jeannin P, Bonnefoy J;		
XX			
DR	WPI; 2001-328651/34.		
DR	N-PSDB; AAF89774.		
XX			
PT	New nucleic acid, expressed in tumours and lymphoid tissue is useful for		
PT	identifying agents for treating tumours and autoimmune disease		
XX			
PS	Claim 10; Page 60-63; 85pp; French.		

Query Match		47.7%;	Score 2712.5;	DB 22;	Length 911;
Best Local Similarity		52.9%;	pred. No. 2.4e-204;		
Matches 482;		Conservative 156;	Mismatches 253;	Indels 21;	Gaps
QY	23	RGSPWPSPAWICCWALAGCQAANAG---	DLPSSSSRPLPPCQEKDYHFEYTECDSSGSR	79	
Db	15	RGRTERRIPR---LWRL----	LLWAGTAFQVTQGTGPELHACKESEYHYEYACDSTGSR	67	
QY	80	WRVAIPNSAVDCSGLPDPVRKECTFCASGEYLEMKNQVCCKGEGTYSLSGSIKFDWE	139		
Db	68	WRVAVPHTPGLCTSLPDPVKTECSFCSCNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEW	127		
QY	140	DELPAGFSNIATFMDTVVGPSPDSRDPGCNNSSWIPRGNYIESNRDDCTVSLIYAVHLKKS	199		
Db	128	DELPHGFASLSANMELDDSAAES-TGNCSTSSKWVPRGDYIASNTDECTATLMYAVNLKQS	186		
QY	200	GYVFEYQYVDNNIFFEFFFIONDQCQEMDTTTTDKWVKLTDNGEWGSHSVMLKSGTNILYW	259		
Db	187	GTVMFEYYPDSSIIFFEFFVNDQCQP-NADDSRWMMKTTTEKG-WEFHSELNRGNVLYW	244		
QY	260	RTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCPKPGTFSNKPGSFNCQVCPRNTYSEKG	319		
Db	245	RTTAFSVWTKVPKPVLRNIAITGVAYTSECFPCPKPGTYADKQGSFCLCPANSYSNKG	304		
QY	320	AKECIRCKDDSQFS--GSSECTERPPCTTKDYFIHTPCDEEGKTQIMYKWIPEKICRED	377		
Db	305	ETSCHQCDPDKYSEKSSSCNVRPACTDKDYFYHTACDANGETQLMYKWKAPKICSED	363		
QY	378	LTDAIRLPPSGEKKDCPPNPGFYNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFE	437		
Db	364	LEGAVKLPASGVKTHCPCPNPGFFKTNNSTCQPCPYGPYSNGS-DCTRCPAGTEPAVGFE	422		
QY	438	YKWNVLPGNMKTSFCFNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLLNLHHPGFKPP	497		
Db	423	YKWNLTPTNMETTIVLSGINFEYKMTGWVAGDHIYTAAGASDNDFMILTLLVVPGRPP	482		
QY	498	TS-MTGATGSELGRITVFETLCSADCVLVFMVDINRKSTNVVESWGVTKEKQAYTHIIF	556		
Db	483	QSVNADTENKEVARITVFETLCSVNCLEYFMVGVNSRNTPTPVETWKSGKQSYTYIIE	542		
QY	557	KNATFTFTWAFQRTNQGDNRRTFNDMVVKIYSITATNAVDGVASSCRACALGSEQSGSSC	616		
Db	543	ENTTSFTWAFQRTTFHEASRYKTYNDVAKIYSINVTNVMNGVASYCRCALEASDVGSSC	602		
QY	617	VPCPPGHYIEKETNOCKECPDFTYLSIHQVYGKEACIPCGPGSKNNODHSVCYSDCFFYH	676		
Db	603	TSCPAGYIDRDSGTCHSCPPNTILKAHPYGVQACVPCGPGTKNNKIHSCLYNDCTFSR	662		
QY	677	EKENQILHYDFSNLSSVGLMNGPSFTSKGTKYFHFFNISLCGHEGKKMALCTNNITDET	736		
Db	663	NTPTRTFNYNFSALANTVTLAGGSFTSKGLKYFHFTLSLCGNQGRKMSVCTDNVTDLR	722		
QY	737	VKEIVAGSDDYTNLVGAFCVQCSTIIPSESKGFRAALSSQSIILADTFIGTVTETLKNIN	796		
Db	723	IPE--GESGFKSITAYVCQAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGIT	779		
QY	797	IKEDMPVPVTSQIPDVHFFYKSSSTATTSINGRSTAVKMRNPTKSGAGVISVPKCPAG	856		
Db	780	SPAEFLHLESGLIPDIVIFFYRSNDVTQSCSSGRSTTIRVRCSPQKTPVPGSLLLPGTCSDG	839		
QY	857	TCDGCTFYFLWESAEACPLTEHDFHEIEGACKRGFQETLYVWNEPKWCIKIGISLPEKKL	916		
Db	840	TCDGCNHFHFWESAAACPLCSVADYRAIVSSCVAGIQKTTYVWREPKLCGGISLPEQVR	899		
QY	917	ATCETVDFWLKV 928			
Db	900	TICKTIDFWLKV 911			

RESULT 8

AAB83851

ID AAB83851 standard; Protein: 870 AA.



XX AAB83851;  
DT 23-JUL-2001 (first entry)  
XX Amino acid sequence of a human protein expressed in tumour cells.  
DE Tumour cell; immunological disease; autoimmune disease; cancer;  
XX infection.  
KW Homo sapiens.  
XX WO200131003-A1.  
OS 03-MAY-2001.  
XX 30-OCT-2000; 2000WO-FR03032.  
PF 29-OCT-1999; 99FR-0013629.  
XX (FABR ) FABRE MEDICAMENT SA PIERRE.  
PA Delneste Y, Magistrelli G, Jeannin P, Bonnefoy J;  
PI New nucleic acid, expressed in tumours and lymphoid tissue is useful for identifying agents for treating tumours and autoimmune disease -  
XX Claim 10; Page 67-70; 85pp; French.  
XX The present sequence represents a human protein expressed in tumour cells. The polynucleotide is useful for screening cDNA/genomic DNA banks and for cloning isolated DNA; identifying mutant forms of the gene that encodes a human protein, where the mutations are associated with abnormal gene expression, or promoters and regulators of the gene, particularly for diagnosis; for recombinant expression of the derived protein; as probes and primers for detection and amplification; and as antisense therapeutics. The tumour expressed protein is useful for raising specific antibodies and to screen agents that modulate its activity, bind to it or interact with it. These agents are potentially useful for treatment or prevention of diseases associated with abnormal expression/activity of the protein, particularly immunological diseases (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic infections.  
XX Sequence 870 AA;  
Query Match 47.6%; Score 2702.5; DB 22; Length 870;  
Best Local Similarity 54.1%; Pred. No. 1.4e-203;  
Matches 473; Conservative 153; Mismatches 237; Indels 11; Gaps 8;  
Qy 58 LPPQCKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKN 117  
Db 5 LHACKESEYHFEYTACDSTGSRWRVAVPHTPGLCTSLPDPVKGTCSFSCNAGEFLDMKD 64  
Qy 118 QVCSKCGEGTYSGLSGIKFDEWDELPAFESNIATFMDTVVVGPSDRPDGCNNSWIPRGN 177  
Db 65 QSCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAES-TGNCCTSSKWVPRGD 123  
Qy 178 YIESNRDDCTVSLIYAVHLKSGYVFFEQYVDNNIFFEQQIQDQCEMDTTTDKWKVL 237  
Db 124 YIASNTDECTATLMYAVNLKQSGTVNFYYPDSIIIEFFVQNDQCP-NADDSRWMT 182  
Qy 238 TDNGEWGSHVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFCKPGT 297  
Db 183 TEKG-WEFHVSVELNRGNVLYWRTTAFSVWTKVPKVLVRNIAITGVAYTSECFCKPGT 241.  
Qy 298 FSNKPGSFNCQVPRNTYSEKAKECIRCKDDSQFS--GSSECTERPPCTTKDYFQIHTP 355  
Db 242 YADKQSSSFCKLCPANSYSNKGETSCHQC-DPDKYSEKSGSSCNVRPACTDKDYFYHTA 300

^ . . ^

Qy 356 CDEEGKTQIMYKWIPIKICREDLTDAILRPPSGEKKDCPPCNPFGYNNSSSCHPCPPGT 415  
Db 301 CDANGETQLMYKWKPKICSEDLGAVKLPASGVKTHCPCPNPGEFFKTNNSTCQPCPYGP 360  
Qy 416 FSDGTKECRPCPAGTEPALGFYKWNVLPGNNKTSFENVGNSKCDGMNWEVAGDHIQS 475  
Db 361 YSNGS-DCTRCPAGTEPAVGFEYKWNVTLPTNMETTVLSGINFEYKGMTGWEVAGDHIYT 419  
Qy 476 GAGGSDNDYLIILNLHIPGFKPPTS-MTGATGSELGRITFVFETLCSADCVLFMVDINRK 534  
Db 420 AAGASDNDMILTLLVPGFRPPQSVMDATENKEVARITFVFETLCSVNCCELYFMVGNSR 479  
Qy 535 STNVVESWGGTKEQAYTHIIFKNATFTFTWAFORTNQGDNRREINDMVKIYSITATNA 594  
Db 480 TNPVETWKGSKGQSYTYIIIEENTTTTFTWAFORTFHEASRKYTNDVAKIYSINVTNV 539  
Qy 595 VDGVASSCRACALGSEQSGSSCVPCPPGHYIEKETNQCKECPDPTLYLSIHQVYKGEACIP 654  
Db 540 MNGVASYCRPCALEASDVSSCTSCPAGYIIDRDSGTSCHSCPPNTILKAHQPYGVQACVP 599  
Qy 655 CGPGSKNNQDHSVCYSDCFYHEKENQILHYDFSLSVSGSLMNGPSFTSKGTKYFHFEN 714  
Db 600 CGPGTKNNKIHSCLYNDCTFSRNTPTRTFNYNFSALANTVTLAGGPSETSKGLKYFHHT 659  
Qy 715 ISLCGHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFCVQCSTIIPSESKGFRALSS 774  
Db 660 LSLCGNQGRKMSVCTDNTDLRIPE---GESGFSKSITAYVCQAVIIPPEVTGYKAGVSS 716  
Qy 775 QSIILADTFIGTVETTLKNNIKEDMFPVPTSQIPDVHFFYKSTATTSCINGRSTAVK 834  
Db 717 QPVSLADRLIGVTTDMTLDGITSPAELFHELSGIPDVIFFYKSNVDVTQSCSSGRSTIR 776  
Qy 835 MRCNPTKSGAGVISVPKCPACTCDGCTFEYFLWESAEACPLCTEHDFHEIEGACKRGFOE 894  
Db 777 VRCSPOKTVPGSLLLPCTGCDGCFNHFHFWESAAACPLCSVADYRAIVSSCVAGIQK 836  
Qy 895 TLYVWNEPKWCIKGISLPEKKLATCETVDFWLKV 928  
Db 837 TTYVWREPKLCSGGISLPEQRTVICKTIDFWLKV 870  
RESULT 9  
AAB48377  
ID AAB48377 standard; Protein; 464 AA.  
XX  
AC AAB48377;  
XX  
DT 20-APR-2001 (first entry)  
XX Human SEC10 protein sequence (clone ID 1795045.0.77).  
DE SECX; cytostatic; gynecological; gene therapy; screening assay; human;  
KW SEC10; chromosomal mapping; forensic biology; cell proliferation; cancer;  
KW cell differentiation; immune associated disorder; gestational disease.  
XX Homo sapiens.  
OS WO200078802-A2;  
XX 28-DEC-2000.  
XX 23-JUN-2000; 2000WO-US17328.  
XX 23-JUN-1999; 99US-0140584.  
PR 20-JUL-1999; 99US-0144722.  
PR 16-SEP-1999; 99US-0154520.  
PR 22-JUN-2000; 2000US-0604286.  
XX (CURA-) CURAGEN CORP.  
PA Shimkets RA, Fernandes E, Vernet C, Yang M, Boldog FL;  
XX Herrmann JL;  
PI











Db 106 TEPAVGFEYKWWNTLPTNMETTVLSGINFEYKGMTGWEVAGDHIYTAAGASDNDFMILT 165  
QY 490 HIPGFKPPTS-MTGATGSELGRITFEETLCSADCVLVYFMVDIN 532  
Db 166 VWPGRPPQSVMDTENKEVARITFVFETLCSVNCELYFMVGVN 209

Search completed: May 12, 2003, 13:18:55  
Job time : 53.5437 secs





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:07:53 ; Search time 22.1915 Seconds  
(without alignments)  
1361.665 Million cell updates/sec

Title: US-10-073-333A-4  
Perfect score: 5681  
Sequence: 1 MLFRARGPVRGRGWGRPAEA.....KEKEDHFESVQLKTSRSPNI 1027

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 294229222 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	165	2.9	3111	2 US-08-460-309-4	Sequence 4, Appli
2	165	2.9	3111	2 US-08-125-077-4	Sequence 4, Appli
3	162	2.9	1193	2 US-08-400-159-10	Sequence 10, Appl
4	162	2.9	1193	3 US-08-611-729A-10	Sequence 10, Appl
5	162	2.9	1253	3 US-08-479-722B-4	Sequence 4, Appli
6	157.5	2.8	1219	4 US-08-882-046-5	Sequence 5, Appli
7	155.5	2.7	1251	5 PCT-US95-02251-3	Sequence 3, Appli
8	155.5	2.7	1252	1 US-08-199-780-3	Sequence 3, Appli
9	155.5	2.7	1252	2 US-08-316-650-3	Sequence 3, Appli
10	153	2.7	1940	2 US-08-644-271-30	Sequence 30, Appl
11	153	2.7	1940	4 US-09-077-955-34	Sequence 34, Appl
12	151	2.7	2594	4 US-08-718-388-7	Sequence 7, Appli
13	151	2.7	5405	4 US-08-718-388-9	Sequence 9, Appli
14	149.5	2.6	2523	1 US-08-185-432-18	Sequence 18, Appl
15	149.5	2.6	2523	4 US-08-899-232-3	Sequence 3, Appli
16	149	2.6	610	1 US-08-365-470-3	Sequence 3, Appli
17	149	2.6	610	3 US-09-209-668-19	Sequence 19, Appl
18	149	2.6	610	4 US-09-009-490A-89	Sequence 89, Appl
19	148	2.6	2703	1 US-08-185-432-19	Sequence 19, Appl
20	148	2.6	2703	4 US-08-899-232-4	Sequence 4, Appli
21	145	2.6	610	6 5217870-2	Patent No. 5217870
22	143	2.5	1404	2 US-08-400-159-2	Sequence 2, Appli
23	143	2.5	1404	3 US-08-611-729A-2	Sequence 2, Appli
24	137.5	2.4	833	4 US-09-013-895A-5	Sequence 5, Appli
25	137.5	2.4	833	4 US-09-448-868-5	Sequence 5, Appli
26	137.5	2.4	2471	1 US-08-185-432-16	Sequence 16, Appl
27	137.5	2.4	2471	1 US-08-083-590A-19	Sequence 19, Appl

28	137.5	2.4	2471	3 US-08-532-384-19	Sequence 19, Appl
29	137.5	2.4	2471	4 US-08-899-232-1	Sequence 1, Appli
30	137	2.4	1010	4 US-08-882-046-7	Sequence 7, Appli
31	137	2.4	1036	4 US-09-068-740A-6	Sequence 7, Appli
32	137	2.4	1187	4 US-09-068-740A-7	Sequence 7, Appli
33	137	2.4	1208	4 US-09-199-865-1	Sequence 1, Appli
34	137	2.4	1218	2 US-08-400-159-6	Sequence 6, Appli
35	137	2.4	1218	3 US-08-611-729A-6	Sequence 6, Appli
36	137	2.4	1218	4 US-08-882-046-2	Sequence 2, Appli
37	137	2.4	1218	4 US-09-214-278-7	Sequence 7, Appli
38	137	2.4	1218	4 US-09-068-740A-11	Sequence 11, Appl
39	137	2.4	1964	4 US-09-467-997-1	Sequence 1, Appli
40	136.5	2.4	1276	4 US-08-937-236-3	Sequence 3, Appli
41	136.5	2.4	1291	4 US-08-569-214-3	Sequence 3, Appli
42	136.5	2.4	1291	4 US-08-937-236-2	Sequence 2, Appli
43	136.5	2.4	1295	4 US-08-569-214-2	Sequence 2, Appli
44	135.5	2.4	810	2 US-08-820-170A-34	Sequence 34, Appl
45	135.5	2.4	810	3 US-09-055-699-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1  
US-08-460-309-4  
; Sequence 4, Application US/08460309  
; Patent No. 5837496  
; GENERAL INFORMATION:  
; APPLICANT: Engvall, Eva  
; APPLICANT: Leivo, Ilmo  
; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin  
; TITLE OF INVENTION: Fragments and Uses Thereof  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/460,309  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/125,077  
; FILING DATE: 22-SEP-1993  
; APPLICATION NUMBER: US PCT/US 94/10730  
; FILING DATE: 21-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/472,319  
; FILING DATE: 30-JAN-1990  
; APPLICATION NUMBER: US 07/919,951  
; FILING DATE: 27-JUL-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LA 9721  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3111 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-460-309-4



Db 654 LLKEESFTIHGTHFPVRKKEFTVLANKRVLLOITYSF-----GMD-----AIFRLSS 702

QY 589 ITATNAV-----DG-VASSCRACALGSEQSGSSCVPCPPGH-----623

Db 703 VNLESAVSYPTDGSIAAVEVCQPPGYTGSSCESCWPRHRRVNGTIFGGICEPCQCFGH 762

QY 624 --YIEKETNOCKECPPTLYLSIHQVYGKEACIPCGRP-----SKNNQDHSVYSDCFFY 675

Db 763 AESCDDVTGECLNCKDHT-----GGPYCDKCLPGFYGEPTKGTSEDCQPCACPLNIP 814

QY 676 HEKENQILHYDFSNLSSVGLMNGPSFTSKGTKYFHFFNISLC--GHEGKKMALCT---- 729

Db 815 SNNFSPATCHLD-----RSLGLICDG-----CPVGYTGPRCERCAEGYF 852

QY 730 -----NNITDFTVKEIVAGSDDDYTNLVGA-FVCQSTIIPSESKGFRAALS 773

Db 853 QPSVPGGSCQPCQCNLDNDFS---IPGSCD--SLSGSCLICKP-----GTTG 895

QY 774 QSIIILADTFIGVTVEITLKNINIKEDMFPVPTSQIPDVHFFYKSSSTATTSCTINGRSTAV 833

Db 896 RYCELCADGYFGDAVDA--KNCQ-----PCRCNAGGSFSEVCHSQTGQCEC---RANVQ 944

QY 834 KMRCNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEHDFHEIEGACKRGFQ 893

Db 945 GORCD-----KCKAGT-----FGLQARGCVPCNCNSFGSKSFDCESGQ 984

QY 894 ETLYVWNEPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAFATVLLVALTCYFWKKN-- 951

Db 985 -----CWCQPGVT--GKKCDRCAHGYNFQEGS-----CTACECSHLGNCD 1024

QY 952 -----QKLEYKYSKLV-----MTTNSKECELPAADS-----CAIMEGEDN 986

Db 1025 PKTGRICPPNTIGEKSKCAPNTWGHISITTGCKACNCSTVGSLDFOCNVNTGQCN 1080

RESULT 3

US-08-400-159-10

; Sequence 10, Application US/08400159

; Patent No. 5869282

; GENERAL INFORMATION:

; APPLICANT: Ish-Horowitz, David

; APPLICANT: Henrique, Domingos M.P.

; APPLICANT: Lewis, Julian H.

; APPLICANT: Myat, Anna M.

; APPLICANT: Fleming, Robert J.

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Mann, Robert S.

; APPLICANT: Gray, Grace E.

; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE

; TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC Compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/400,159

; FILING DATE: 07-MAR-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-029

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-9741/8864

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1193 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-400-159-10

Query Match 2.9%; Score 162; DB 2; Length 1193;

Best Local Similarity 18.6%; Pred. No. 1.3e-05;

Matches 237; Conservative 116; Mismatches 420; Indels 498; Gaps 65;

QY 91 CSGLPDPVRKECTFSCASGEYLEMKNQVCSCGEGTYSLSGSIKFDWDELPAQFSNIA 150

Db 29 CDGTRNP-GDKKCT-----RDECDTYFKVCLKEYQSRVTAG-----PCSFSGKS 72

QY 151 TFMDTVVGPDSRDPDGCNNSSWIIPRGNYIESNRDDCTVSLIYAVHLKKSGYVF-FEYQYV 209

Db 73 T---PVIG-----GN-----TFNLKYSRNNNEKNRIVIPFTFAWP 103

QY 210 DN-NIFFEFFIQDQCQEMDTTQKVKLTDNGEWSHVMKLSGNTILYWRT----TGI 264

Db 104 RSYTLLEAWDYNDNSTNPDRIEK-----ASHSGMINPSRQ---WOTLKHTGA 150

QY 265 LMGSKAVKPVLVKNITIEGVAYTSECFP-CKPGTFSNKPQSFNCQVCPRNTYSEKGAKEC 323

Db 151 AHFEYQIR-----VTCAEHYGFGCNKFCRP-----RDDFFTHHTCDQN-----GNKTC 194

QY 324 IRCKDDSQFSGSSECTE---RPPCTTKDYFQIHTPCDEEGTKQIMYKW-----368

Db 195 L-----EGWTG-PECNKAICRQCSPK-----HGSCVTPGECRCQYGWQGYCDKCIPHP 243

QY 369 -----IEP-----KICREDLTDAILRPP-----SGEKKDCPPCNPFGY 401

Db 244 GCVHGTCTIEPWQCLCETNWWGQLCDKDLNYCGTHPPCLNGGTCSTGPDKYQCSCEGYS 303

QY 402 NNG-----SSSCH-----PCPPG---TFSDGTKECRPCPAG-----429

Db 304 GQNCETAEHACLSDPCCHNGGSCLETSTGFECVCAPGWAGPTCTDNIDDCSPNPGHGGTC 363

QY 430 TEPALGFY---KWNVNL-----PGNMKTSFCN-VGNSKCDGMNGWE-----467

Db 364 QDLVDGFKCICPPQWTGKTCLDANECEGKPCVNANSRNLIGSYDCITGWSGHNCDI 423

QY 468 -----VAGDHIQSGA-----GGSDNDYL---485

Db 424 NINDCRGQCQNGGSCRDLVNGYRCICSPGYAGDHCEKDINECASNPCMNGGHCQDEINGF 483

QY 486 -----ILNLHIPGFKPPTSMTGATGSELGRITFV-----FE-----516

Db 484 QCLCPAGFSGNLCQLDIDYCEPNPCQNGAQCENLAMDYFCNCPEDYEGKNCSHLKDHCRT 543

QY 517 TLCSA--DCVLYFMVD-----INRKSTNVVESWGGTKEKQAYTHIIFKNATFTTWAQQR 569

Db 544 TPCEVIDSCTVAVASNSTPEGVRYISSNVCGPHGKCKSQAGGKFTCECNKGFTGTYCHEN 603

QY 570 TNQGQDNRRFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGHYIEKET 629

Db 604 INDCESNP-----CKNGGTCIDGVNSYKCICSDGWE-----GTYCETNI 642

QY 630 NQCKECPDPTLYLSIHQVYGKEACIPCGRP-----SKNNQ-DHSVCYSDCFYHEKENQ 681

Db 643 NDCSKNPCHNGGTCRDLVNDDFEC-ECKNGWKGKTCHSRDSQCDTEATCNGGTCYDEGDT- 700

QY 682 ILHYDFSNLSSVG-----SLMNGPSFTSKGTKYFHFFNISL 717

Db 701 -----FKCMCPAGWEGATNIARNSSCLPNPCHNGGTCVVSGDSFT-----CV 743

QY 718 CGHEGKKMALCTNNTIDFTVKEI-----VAGSDDYTNLVGAFFVQ 757

Db 744 C-KEGWEGPTCTQNTNDCSPHPCYNSTGTCVDGDNWYRCECAPGFAGPDCRINI---NECQ 799



QY 758 ST-----IIPSESKGF-----RAALSSQSIIILADTFIGVTVETTLKNINIKEDMPV 804  
Db 800 SSPCAFGATCVDEINGYRCICPPGRSGGCQEVTRPCFTSIRV-----843  
QY 805 PTSQIPDVHFFYKSSSTATTSINGRSTAVKMRCP-----TKSGAGVISVPSK-- 852  
Db 844 ---MPD-GAKWDDDCNTCQLNGKVTCSKVWCGPRPCIIHAKGHNECPAGHACVPVKED 898  
QY 853 -----CPA-GTC-----DGC---TFYFLWESAEACPLCTEHDFFE 883  
Db 899 HCFTHPCAAAGECWPSNQPVKTKCNSDSYYQDNCANITFTFNKEMM-APGLTTEHCSE 957  
QY 884 IE--GACKRGFOE-TLYVWNEPKWCI-----KGISLPEKK 915  
Db 958 LRNLNLKNVSAEYSIITCEPSHLANNEIHVAISAEDIGEDENPIKEITDKIIDLVSKR 1017  
QY 916 -----LATCET-----VDFWLKVGAGVGAFPTAVLLVALTC-----YFW-----K 949  
Db 1018 DGNNTLIAAAVAEVRVQRRPVKNKTDFL-----VPLLSSVLTVAWICCLVTVFYWCIOKR 1071  
QY 950 KNQKLEYKYSKLVMTTNSKECELPAAADSCAIMEGDNBEVWYSNKQSLGKLSLATKE 1009  
Db 1072 RKQSSHTHTASDNTTNNVREQLNQIKNPIEKHGANTVPIKDYEKNKNSKIAKIRTHNSEV 1131  
QY 1010 KEDHFESVLK 1020  
Db 1132 EEDDMDKHQK 1142

RESULT 4

US-08-611-729A-10  
; Sequence 10, Application US/08611729A  
; Patent No. 6004924  
; GENERAL INFORMATION:  
; APPLICANT: Ish-Horowicz, David  
; APPLICANT: Henrique, Domingos M.P.  
; APPLICANT: Lewis, Julian H.  
; APPLICANT: Myat, Anna M.  
; APPLICANT: Fleming, Robert J.  
; APPLICANT: Artavanis-Tsakonas, Spyridon  
; APPLICANT: Mann, Robert S.  
; APPLICANT: Gray, Grace E.  
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE  
; TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/611,729A  
; FILING DATE: 06-MAR-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 7326-037  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1193 amino acids

; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-611-729A-10  
Query Match 2.9%; Score 162; DB 3; Length 1193;  
Best Local Similarity 18.6%; Pred. No. 1.3e-05;  
Matches 237; Conservative 116; Mismatches 420; Indels 498; Gaps 65;  
QY 91 CSGLPDPVRGKECTFSCASGEYLEMKNQVCSKCGEGYSILSGIKFDEWDELPAFNSIA 150  
Db 29 CDGTRNP-GDKKCT-----RDECDTYFKVCLKEYQSRVTAGG-----PCSFSGSKS 72  
QY 151 TFMDTVVGPDSRDPDGCNNSSWIPRGNIESNRDDCTVSLIYAVHLKKSGYVF-FEYQYV 209  
Db 73 T---PVIG-----GN-----TFNLKYSRNEKNRIVIPFIFAMP 103  
QY 210 DN-NIFFEFFIIONDQCQEMDTTDDKWKVLTLDNGEWGSHSVMLKSGTNILYWRT----TGI 264  
Db 104 RSYTLLEAWDYNDSNTPDRIEK-----ASHSGMINPSRQ---WQTLKHINTGA 150  
QY 265 LMGSKAVKPVLVKNITIEGVAYTSECFP-CKPGTFSNKPGSFNCQVCPRNTYSEKAKEC 323  
Db 151 AHFEYQIR-----VTCAEHYGFGCNKFCRP-----RDDEFTHTCDQN-----GNKTC 194  
QY 324 IRCKDDSQFSGSECTE---RPPCTTKDYQIHPTPCDEGKTQIMYKW-----368  
Db 195 L-----EGWTG-PECNKAICRQGCSPK-----HGSCVTPCECRQYGVQGWQCYCDKCTPHIP 243  
QY 369 -----IEP-----KICREDLTDAIRLPP-----SGEKKDCPPCNPFGY 401  
Db 244 GCVHGTCTIEPWQCLCETNWSGQLCDKDLNYCGTHPPCLNGGTCNTGPDKYQCSCPEGYS 303  
QY 402 NNG-----SSSCH-----PCPPG-----TFSDGTKECRPCPAG-----429  
Db 304 QNCEIAEHAELSDPCCHNGSGCLETSTGFECVCAPGWAGTCTCTDNIDDCSPNCPGHHGTC 363  
QY 430 TEPALGFY-----KWNVL-----PGNMKTSFEN-VGNSKCDGMNGWE-----467  
Db 364 QDLVDGFKCICPPQWTGKTOLDANECEGKPCNVANSRNLIGSYCDCTIGWSGHNCDI 423  
QY 468 -----VAGDHIQSGA-----GGSDNDYL---485  
Db 424 NINDCRGQCONGGSCRDVLNRYRCISPGYAGDHCEKDINECASNPMMNGGHCCQDEINGF 483  
QY 486 -----ILNLHIPGFKPPTSMTGATGSELGRITFV-----FE-----516  
Db 484 QCLCPAGFSGNLCQLDIDYCEPNPCQNGAQCFNLAMDYFCNCPEDYEGKNCSHLKDHCRT 543  
QY 517 TLCSA--DCVLYFEMVD-----INRKSTNVVESWGTTKEKQAYTHIIFKNATFTTWAQOR 569  
Db 544 TPCEVIDSCTVAVASNSTPEGVYLISSNVCGPHGKCKSQAGGKFTCECNKGFTGTYYCHEN 603  
QY 570 TNOQDNRRRFINDMVKIYSITATNAVDGVASSCRACALGSESGSSCVPCPPGHYIEKET 629  
Db 604 INDCESNP-----CKNGGTCIDGVNSYKICISDGE-----GTyceTNI 642  
QY 630 NOCKECPDPTYLSIHQVYGKEACIPCGPG-----SKNQ--DHSVCYSDCFYHEKENQ 681  
Db 643 NDCSKNPCHNGGTCRDLVNDFFEC-ECKNGWKGTCHSRDSQCDATCNGGTCYDEGDT- 700  
QY 682 ILHYDFSNLSSVG-----SLMNGPSFTSKGTYFHEFNISL 717  
Db 701 -----FKCMCPAGWEGATCNIARNSSCLPNPCHNGGTCVVVSGDSFT-----CV 743  
QY 718 CGHEGKKMALCTNNITDFTVKEI-----VAGSDDYTNLVGAFVCQ 757  
Db 744 C-KEGWEPTCTQNTNDCSPHPCYNSGTCVDGDNWYRCECAPGFAGPDCRINI---NECQ 799  
QY 758 ST-----IIPSESKGF-----RAALSSQSIIILADTFIGVTVETTLKNINIKEDMPV 804  
Db 800 SSPCAFGATCVDEINGYRCICPPGRSGGCQEVTRPCFTSIRV-----843

QY 805 PTSQIPDVHFFYKSSSTATSCINGRSTAVKMRCP-----TKSGAGVISVPSK-- 852  
Db 844 ----MPD-GAKWDDDCNTCQCLNGKVTCSKVWCGPRPCIIHAKGHNECPAGHACVPVKED 898  
QY 853 -----CPA-GTC-----DGC---TFYFLWESAEACPLCTEHDHFHE 883  
Db 899 HCFTHPCAAGGECWPSNQPVKTKCNSDSYYQDNCANITFTFNKEMM-APGLTTEHICSE 957  
QY 884 IE--GACKRGFQE-TLYVWNEPKWCI-----KGISLPEKK 915  
Db 958 LRNLNLKNVSAEYSIVITCEPSHLANNEIHVAISAEDIGEDENPIKEITDKIIDLVSKR 1017  
QY 916 -----LATCET-----VDFWLKVGAGVGAFATVLLVALTC-----YFW-----K 949  
Db 1018 DGNNTLIAAUAVERVORRPVKNKTDFL-----VPLLSSVLTVAWICCLVTVFYWCIOKR 1071  
QY 950 KNQKLEYKYSKLVMTNTSKECELPAAADSCAIMEGEDNEEVVYSNKQSLGLKLSLATKE 1009  
Db 1072 RKQSSHHTASDDNTNNVREQLNQIKNPIEKHGANTVPIKDYENKNKSKIAKIRTHNSEV 1131  
QY 1010 KEDHFESVOLK 1020  
Db 1132 EEDDMDKHQK 1142

RESULT 5  
US-08-479-722B-4  
; Sequence 4, Application US/08479722B  
; Patent No. 6074840  
; GENERAL INFORMATION:  
; APPLICANT: Bonadio, Jeffrey  
; APPLICANT: Yin, Wushan  
; TITLE OF INVENTION: LATENT TGF( BINDING PROTEIN (LTBP)  
; TITLE OF INVENTION: GENES, COMPOSITIONS AND METHODS  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Williams, Morgan & Amerson  
; STREET: 7676 Hillmont, Suite 250  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77040  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/479,722B  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US95/02251  
; FILING DATE: 21-FEB-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/316,650  
; FILING DATE: 30-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/199,780  
; FILING DATE: 18-FEB-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fussey, Shelley P.M.  
; REGISTRATION NUMBER: 39,458  
; REFERENCE/DOCKET NUMBER: 4100.000500/FUS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (713) 934-7000  
; TELEFAX: (713) 934-7011  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1253 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-479-722B-4  
Query Match 2.9%; Score 162; DB 3; Length 1253;  
Best Local Similarity 20.3%; Pred. No. 1.4e-05;  
Matches 150; Conservative 67; Mismatches 246; Indels 276; Gaps 48;  
QY 289 ECFPCPKP--GTFNPKGFSNQCVCPR--NTYSEKGAKECIRCKDDSQFSGSSECTERPPC 344  
Db 601 EAEPGCGPGKIGCMNTGGSYNCH--CNRGYRLHVGAGGRSCV-----DLNECAKPHLC 650  
QY 345 TTKDY---FQIHTPCDEEGKTIQIMY-----KWIEPKICREDLTDAIRLP---PSGEKKDC 393  
Db 651 GDGGFCINFPGHYKCN-----CYPGYRLKASRPPIIC-EDI-DECRDPSTCPDGKCKENK 701  
QY 394 PPCNPGFYNNNGSSSCHPCPPGTFTSDGTKECR---PCPAGTEPALGFEYKWNVLPGNMKT 450  
Db 702 P-----GSPKCIACQPGYRSQGGGACRDVNECSEGTPCSPG---WCENLPGSYRC 748  
QY 451 SC-----FNVGNSKC---DGMNGWEVAGDHIQSGAGGSDNDYLIILNLHIPGFKPPTMTG 502  
Db 749 TCAQGIRTRTGRLSCIDVDCEAGKVCQDGTCTNTPGSFQCQCLSGYHL-----797  
QY 503 ATGSELGRITFV---FETLC-SADCVLYFMVDINRKSTNVVESW-----GGTK 546  
Db 798 --SRDRSRCEDIDECDFPAACIGGDCI-----NTNGSYRCLCLPLGHRHLVGGRK 843  
QY 547 -----EKQAYTHIIFKNATFTT---WAFQRTNQGDNR--FIN- 581  
Db 844 CKKIDIDECSDPGLCLPHACENLQGSYVCVDEGFTLTQDQHGCEEVEQPHHKCYLNF 903  
QY 582 -DMVKIYSITATNAVDGVASSCRACALGSEQSGSSC--VPC-----PPGHYIE 626  
Db 904 DDTVFCDSVLATNVQT---QEC-CCSLGAGW-GDHCEIYPCPVYSSAEFHSVLPDGKRLH 958  
QY 627 KETNQCKECPDPTYLSIHQ--VYGKEACI--PCGPGSKNNQDHSVCYSDCFFYHEKENQI 682  
Db 959 SGQQHCELCIP-AHRDIDECILFGAEICKEGKC---VNTQPGYECYCKQGFY-----1006  
QY 683 LHYDFSNLSSVGLMNGPSTSKGTKYPHFFNISLCGHEGKKMALCTNNITDFTVKEIVA 742  
Db 1007 --YDGNLLECV-----DVDECLDES-----CRNGVCE-----1032  
QY 743 GSDDYTNLVGAFVCQSTIIPSESKGFRAALSSQSIILADTFIGVTETTLKNINIKEDMF 802  
Db 1033 -----NTRGGYRCAC-PPAEYSAPAQ---QCLI-----PERW 1061  
QY 803 PVPTSQIPDVHFFYKSSSTATSCING-----RSTAVKMRCPN-TKSGAGVI 847  
Db 1062 STPQRDVKCA---GASEERTACVWGPWAGPALTFDDCCCRQPRLGTQCRPCPPRGTG-- 1115  
QY 848 SVPSKCPAGTCDGCTFYFLWESAEACPLC-----TEHDFHEIEGACKRGFQETLYVWNE 901  
Db 1116 ---SQCPSTQSESNF---WDTs---PLLLGKSPRDESDSESDSECR-----1154  
QY 902 PKWCIKGISLPEKKLATCE 920  
Db 1155 ---CVSGRCVPRPGGAVCE 1170

RESULT 6  
US-08-882-046-5  
; Sequence 5, Application US/08882046  
; Patent No. 6136952  
; GENERAL INFORMATION:  
; APPLICANT: Li, Linheng  
; APPLICANT: Hood, Leroy  
; APPLICANT: Krantz, Ian D.  
; APPLICANT: Spinner, Nancy B.  
; TITLE OF INVENTION: Human Jagged Polypeptide, Encoding  
; TITLE OF INVENTION: Nucleic Acids and Methods of Use  
; NUMBER OF SEQUENCES: 110  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP







QY 743 GSDDYTNLVGAFCVQSTIIPSESKGFRAALSSQSIIILADTFIGVTVETTLKNINIKEDMF 802  
Db 1037 -----LPCACT-PPAEYSPAQA-----QCLI-----PERW 1060  
QY 803 PVPTSQIPDVHFFYKSSSTATTSING-----RSTAVKMRGNP-TKSGAGVI 847  
Db 1061 STPQRDVKA---GASERTACVWGPWAGPALTFDDCCCRQPRLTQCRPCPPRG TG-- 1114  
QY 848 SVPSKCPAGTCDGCTFYFLWESAECPLC-----TEHDFHEIEGACKRGFOETLYVWNE 901  
Db 1115 ---SQCPSTQSESNF---WDTs---PLLLGKSPRDESDSEDSDECR----- 1153  
QY 902 PKWCIKGISLPEKKLATCE 920  
Db 1154 ---CVSGPCVPRPGGAVCE 1169

RESULT 9  
US-08-316-650-3  
; Sequence 3, Application US/08316650  
; Patent No. 5942496  
; GENERAL INFORMATION:  
; APPLICANT: Bonadio, Jeffrey  
; APPLICANT: Roesler, Blake J.  
; APPLICANT: Goldstein, Steven A.  
; APPLICANT: Lin, Wushan  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS  
; TITLE OF INVENTION: FOR STIMULATING BONE CELLS  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/316,650  
; FILING DATE: 30-SEP-1994  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/199,780  
; FILING DATE: 30-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, David L.  
; REGISTRATION NUMBER: 32,165  
; REFERENCE/DOCKET NUMBER: UMIC:008  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (713) 789-2679  
; TELEX: 79-0924  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1252 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-316-650-3

Query Match 2.7%; Score 155.5; DB 2; Length 1252;  
Best Local Similarity 20.2%; Pred. NO. 5.5e-05;  
Matches 149; Conservative 68; Mismatches 245; Indels 277; Gaps 48;  
QY 289 ECFPCPKP--GTFSNKPQSFNCQVCP--NTYSEKGAKECIRCKDDSQFSGSSECTERPPC 344  
Db 601 EAEPCGPGKICMNTGGSYNCH-CNRGYRLHVGAGRSCV-----DLNECAKPHLC 650  
QY 345 TTKDY---FQIHTPCDECKTQIMY-----KWIEPKICREDLTDAILRP---PSGEKKDC 393

Db 651 GDGFCINFPGHYKCN-----CYPGYRLKASRPPIK-EDI-DECRDPSTCPDGKCIENK 701  
QY 394 PPCNPGFYNNNGSSSCHPCPPPGTFSDGTKECR---PCPAGTERPALGFEYKWNVLPGNMTK 450  
Db 702 P-----GSFKCIAQCPGYRSGGGACRDVNECSEGTPCS PG---WCEKLPGSYRC 748  
QY 451 SC-----FNVGNSKC---DGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTMTG 502  
Db 749 TCAQGIKTRTGRSLCIDVDDCEAGKVCQDGICITNTPGSFQOCLSGYHL----- 797  
QY 503 ATGSELGRITFV---FETLC-SADCVLYFMVDINRKSTNVVSW-----GGTK 546  
Db 798 --SRDRSRCEIDECDFPAACIGGDCI-----NTNGSYRCLCPLGHRVVGGRK 843  
QY 547 -----EKQAYTHIIFKNATFTT---WAFQRTNQGDNR--FIN- 581  
Db 844 CKKDIDECSDQDPLCLPHACENLQSYVVCVDEGFTLTQDQHGCEVEQPHHKKECYLNF 903  
QY 582 -DMVKIYSTATNAVDGVASSCRACALGSEQSGSSC--VPC-----PPGHYIE 626  
Db 904 DDTVFCDVLTATNVQT---QEC-CCSLGAGW-GDHCEIYPCPVYSSAEFHSVLPDGLKRLH 958  
QY 627 KETNQCKECPDPTLYLSIHQ--VYGKEACI--PCGPGSKNNQDHSVCYSDCFFYHEKENQI 682  
Db 959 SQQOHCCLCIP-AHRDIDECILFGAEICKEGKC---VNSQPGVECYCKQGFY----- 1006  
QY 683 LHYDFSNLSSVGLMNGPSFTSKGTIFYHFFNISLCGHEGKMMALCTNNITDFTVKEIVA 742  
Db 1007 --YDGNLLEC-----DVDECLDES-----CRNGVCENTWR--- 1036  
QY 743 GSDDYTNLVGAFCVQSTIIPSESKGFRAALSSQSIIILADTFIGVTVETTLKNINIKEDMF 802  
Db 1037 -----LPCACT-PPAEYSPAQA-----QCLI-----PERW 1060  
QY 803 PVPTSQIPDVHFFYKSSSTATTSING-----RSTAVKMRGNP-TKSGAGVI 847  
Db 1061 STPQRDVKA---GASERTACVWGPWAGPALTFDDCCCRQPRLTQCRPCPPRG TG-- 1114  
QY 848 SVPSKCPAGTCDGCTFYFLWESAECPLC-----TEHDFHEIEGACKRGFOETLYVWNE 901  
Db 1115 ---SQCPSTQSESNF---WDTs---PLLLGKSPRDESDSEDSDECR----- 1153  
QY 902 PKWCIKGISLPEKKLATCE 920  
Db 1154 ---CVSGPCVPRPGGAVCE 1169

RESULT 10  
US-08-644-271-30  
; Sequence 30, Application US/08644271  
; Patent No. 5814478  
; GENERAL INFORMATION:  
; APPLICANT: Valenzuela, et al.  
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS  
; TITLE OF INVENTION: AND LIGANDS  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
; STREET: 777 Old Saw Mill Road  
; CITY: Tarrytown  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10591  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/644,271  
; FILING DATE: 10-MAY-1996  
; CLASSIFICATION: 435

[illegible]

Db 387 RVCSGIYDPVCGSD-----GVYGSVCELESMAC-----TLGREIQVARRG---- 427

QY 484 YLILNLHIPGFKPPTMTGAT--GS-----ELGRITVFETLCSADCVLYFMVDINRKSTN 537

Db 428 -----PCDPCGQCRCFGLCEVETGR-----CVCPSCV-----ESAQP 460

QY 538 VVESWGGTKEKQAYTHIIFKNATFTTWAQRTNQODNRRFINDMVKIYSITATNAVDG 597

Db 461 VCGSDGHTYASECELHV-----HACTHQISLYVASA-----G 492

QY 598 VASSC--RACALGSEQSSCVPCPPGPHYIEKETNOCKECPD-----TVLSI----- 643

Db 493 HCOTCGEKVCTFGAVCSAGCV-CP-----RCEHPPPGPVCGSDGVTYLSACELR 541

QY 644 ----HQVYGKEA---CIP--CGPGSKNNQDHSVCYSD-C-----FFYHEKENQILHYD 686

Db 542 EAACQQQVQIEEAHAGPCEPAECGSGSGEDDECEQELCRQGGIWDSEDSGPCVCD 601

QY 687 FSNLSSVGLMNGPSFTSKGTYKFFHFNISLCGHEGKMACTNNITDFTVKEIVAGSDD 746

Db 602 F-----SCQSVPRSP-----VCGSDGVYTG-----TECDLKKARCESQQ 635

QY 747 YTNLVGAFVCQS-TIIPSESKGFRAALSSQSIILADTFIGVTVETTLKNINIKEDMFPVP 805

Db 636 ELYVAAQACRGPTLAP-----LLPVA 657

QY 806 TSQIPDVHFFYKSTATTSCINGRSTAVKMRNPTKSGAGVISVPSKCP-----AGTC 858

Db 658 -----FPHCAQTPYGCCODNFTAAQ-----GVGLAGCPSTCHCNPHGYSGTC 700

QY 859 DGCT-----FY-----FLWESAECPLCT-----EHDFFEIEGA 887

Db 701 DPATGQSCRPGVGLRCRCEPGFWNFRGIVTDGHSRGCTPCSDPRGAVRDDCEQMTGL 760

QY 888 CKRGFOETLYWNEPKWCIKIGISLPEKKLATCETVDFWLKVAGAGVGAFTAVLLVALTCYF 947

Db 761 CS-----CRPGVAGP--KCGQCPDGQVLHGLGCEADPMTPTVTCVEIHCBF 803

QY 948 WKKNQKLEYKYSKLVMTTNSKEC-----ELPAADSCAIMEGEDNEEEVYSNKSLLGLKL 1002

Db 804 G-----ASCVEKAGFAQCICPTLTCPLEANSTKVC-GSDG---VTYGNE-----CQL 845

QY 1003 KSLATKEKD 1012

Db 846 KAIACRQRULD 855

RESULT 12

US-08-718-388-7

; Sequence 7, Application US/08718388

; Patent No. 6271362

; GENERAL INFORMATION:

; APPLICANT: MORIKAWA, MINORU

; APPLICANT: HARADA, NAOKI

; TITLE OF INVENTION: GENE ENCODING IgG Fc REGION-BINDING

; TITLE OF INVENTION: PROTEIN

; NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH

; STREET: PO BOX 747

; CITY: FALLS CHURCH

; STATE: VA

; COUNTRY: USA

; ZIP: 22040-0747

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/718,388

; FILING DATE:

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: MURPHY JR, GERALD M

; REGISTRATION NUMBER: 28,977

; REFERENCE/DOCKET NUMBER: 0230-111

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 205-8000

; TELEFAX: (703) 205-8050

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2594 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-718-388-7

Query Match 2.7%; Score 151; DB 4; Length 2594;

Best Local Similarity 17.5%; Pred. No. 0.00047;

Matches 202; Conservative 103; Mismatches 346; Indels 502; Gaps 62;

QY 18 AEAPRRGRSPWPATWICCWALAGCAAAGDLPSS-----SRPLPPCQEKDYHFEYT 71

Db 765 AAAPSNCSGRPCVEGVC---LPGFVASGGACVPASSCGCTFGQLQAPGQEV----- 814

QY 72 ECDSSGSRWRVAIPNSAVDCSGLPDV--RGKECTFSCASGEYLEMKNQV----- 119

Db 815 -----WADLCQRCTCNGATHQVTCRDQK---SCPAGERCSVQNGLLGCGYDPRFGT 863

QY 120 CSKCGEGTYSLGSIGIKFDEWDELPAFESNIATFMDTVVGPSPDSRPGCNSNSWIPRGNYI 179

Db 864 CQSGDPHYVSFDRRED-----FMGTCTYL--LVG-----SCGQNAALPAFKVL 906

QY 180 ESN--RDDCTVSLIYAVHLKKSGY--VFFEYQ---YVDNNI--FFEFFIQNDQCQEMDIT 230

Db 907 VENEHRGSQTVSYTRAVRVEARGVKAVRREYPGQVLDVDDVLYLPFQAADGQVQVFRQG 966

QY 231 TDRWKV---LTDNGEWGSH--SVMLKSGTNILYWRITTGILMGSKAVKPVL-----VKNIT 280

Db 967 RDAVVRTDFGLTVTYDMNARVTAKVPSSYAEALCGLCGNFNGDPADDLALRGGQAANAL 1026

QY 281 IEGVAYTSECFCKPGTFESNKPSENCQVCP-----RNT 314

Db 1027 AFGNSWQEE---TRPGCGGATEPGD-----CPKLDLSIAQQLQSKNECGILADPKGPFREC 1078

QY 315 YSE---KGA-KECI--RCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEGKTQIMYKW 368

Db 1079 HSKLDPQGAVRDCVYDRCLLPQO-SG-----PLCDALATYA--AACQAAGAT--VHPW 1126

QY 369 IEPKICREDLTDAIRLPPSGEKKDCP-----PCNPGFYNNG 404

Db 1127 RSEELC-----PLSCPPIISHYEAACSYGCLSCGDLVPVGGCGSECHCGVCDEGFALSG 1180

QY 405 SSSCHP-----C-----PPG-TF-----SDGTKECRPCPAGTEPALGFEY 438

Db 1181 -ESCLPLASCGCVHQGTYPHPGQTFYPPGCDLSLCHCQEGGLVSCSSSCGPHEA----- 1234

QY 439 KWNVNLPGNMKMTSCFNVGNKSCDGMNGWEVAGDHIQSGAGGSDNDYLIILNLHIPGFKPPT 498

Db 1235 ---CQPSGSGSLGCVAVGSSTC-----QASGD-----PHY---T 1261

QY 499 SMTGATCSELGRITFVFETLCSADCVLYFMVDINRKSTNVVSWGGTKEKQAYTHIL-FK 557

Db 1262 TFDGRREDFMGTCVYVLAQTCGTRPGLHREFAVLQE---NV--AWG--NGRVSVTRVITVQ 1314

QY 558 NATFTF-----TW-----AFQRTNQGD-----NRRFINDMVKIY 587

Db 1315 VANFTLRLEQRQWKVTVNGVDMKLPVVLANGQIRASQHGSDVVIETDFGLRVAYDLV--Y 1372

QY 588 SITATNAVDGVASSCRACA-----LGSEQSG-----SSCV--- 617

Db 1373 YVRVTVPGNYQQMCGLCGNYNGBDKDDFKPNGSQAGNANEFNGSWEEVVPDSPCLPPT 1432

QY 618 PCPPGHIETKETNQCKECPDPTYLSIHQVYKGAICIPCGPGSKNNQDHSVCY----- 669



```

Db 1433 PCPPG---SEDCIPSHKCPPE-----LEKKYQKEEF--CGLLSSPTGPLSSCHKLVDPQGP 1483
QY 670 -SDCFFYHEKENQILHYDFSNLSSVGLMNGPSFTSKGTKYFHHFFNISLCGHEGKKMALC 728
Db 1484 LKDCIF-----DLCLGGGNLSILC 1502
QY 729 TNNITDFTVKEIVAGSDDYTNLVGAFV--CQST---IIPSESKGFRAALSSQSIIILADTF 783
Db 1503 SN-----IHAYVSACQAAGGHVEPWRTETF-----1527
QY 784 IGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSATTSCINGRSTAVKMRCPNPTKSG 843
Db 1528 -----CPMECPPNSHYELCADTCSLGC-----1549
QY 844 AGVISVPSKCPAGTCDG--CTFYFLWESAEACPL-----875
Db 1550 -SALSAPQCQDGCAGCQCDSDGFLYNGQACVPIQQCGCYHNGVYYPEQTVLIDNCRQQ 1608
QY 876 -----CTEHD-----FHEI-----EGACKRG 891
Db 1609 CTCHAGKGMVCQEHSCKPGQVCQPSGGILSCVTKDPCHGVTCRPQETCKEQGGQGVCLPN 1668
QY 892 FQETLYVWNEPKW 904
Db 1669 YEATCWLWGDPHY 1681

RESULT 13
US-08-718-388-9
; Sequence 9, Application US/08718388
; Patent No. 6271362
; GENERAL INFORMATION:
; APPLICANT: MORIKAWA, MINORU
; APPLICANT: HARADA, NAOKI
; TITLE OF INVENTION: GENE ENCODING IgG Fc REGION-BINDING
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,388
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0230-111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5405 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-718-388-9
```

Query Match 2.7%; Score 151; DB 4; Length 5405;  
Best Local Similarity 17.5%; Pred. No. 0.0016;  
Matches 202; Conservative 103; Mismatches 346; Indels 502; Gaps 62;

```

QY 18 AEAPRRGRSPWPSPAWICCWALAGCQAAWAGDLPSS-----SRPLPCQCKDYHFEYT 71
Db 765 AAAPSNCSGRPCVEGCVC--LPGFVASGGACVPASSCGCTFQGLQLAPQEV-----814
QY 72 ECDSSGSRWRVAIPNSAVDCSGLPDPV--RGKECTFSCASGEYLEMKNQV-----119
Db 815 -----WDELQRRCTCNGATHQVTCRDQK---SCPAGERCSVQNGLLGCYPDRFGT 863
QY 120 CSKCGEGTYSLSGKIFDEWDELPAFNSIATFMDTVVGPDSRDPDGCNNSWIPRGNVI 179
Db 864 CQSGDPHYVSFDGRRFD-----FMGTCTYL--LVG-----SCGQNAALPAFRVL 906
QY 180 ESN--RDDCTVSLIYAVHLKSGY---VFFEYQ---YVDNNI--FFEFFIQNDQCQEMDTT 230
Db 907 VENEHRGSQTVSYTRAVRVEARGVKAVRREYPGQVLVDVQLYLPFOAADGQVQVFRQG 966
QY 231 TDKWVK----LTDNGEWGSH-SVMLKSGTNILYWRRTTGILMGSKAVKPVL-----VKNIT 280
Db 967 RDAVVRTDFGLTVTYDWNARVTAKVPSVSYAEALCGLCGNFNGDPADDDLALRGGQAANAL 1026
QY 281 IEGVAYTSECPCKPGTFSNKPGSFNCQVCP-----RNT 314
Db 1027 AFGNSWQEE---TRPGCGATEPGD-----CPKLDLSVAQQLQSKNECILADPKGPFREC 1078
QY 315 YSE--KGA-KECI--RCKDDSQFSGSSECTERPPCTTKDYFQIHPTCDEEGKTQIMYKW 368
Db 1079 HSKLDPOGAVRDCVYDRCLLPQG-SG-----PLCDALATYA--AACQAAGAT--VHPW 1126
QY 369 IEPKICREDLTDAIRLPPSGEKKDCP-----PCNPGFYNNG 404
Db 1127 RSEELC-----PLSCPPHSHYEACSYGCLSCGDLVPVGGCGSECHGCVDEGFALSG 1180
QY 405 SSSCHP-----C-----PPG-TF-----SDGTKECRPCPAGTEPALGFY 438
Db 1181 -ESCLPLASCCGVHQGTYPHPPGQTFYPGCGDSLCHCQEGGLVSCSSCGPHEA-----1234
QY 439 KWNVLPGNMKTSCFNVGNKCDGMNGWEVAGDHIQSGAGSDNDYLILNLHIPGFKPPT 498
Db 1235 ----CQPSGSLGCVAVGSSTC-----QASGD-----PHY---T 1261
QY 499 SMTGATGSELGRITVFETLCSADCVLVFMVDINRKSTNVVESWGSKTKQAYTHII-FK 557
Db 1262 TFDGRRRDEMGTCVYVLAQTCGTRPGLHFAVLQE--NV--AWG--NGRVSVTRVITVQ 1314
QY 558 NATETF-----TW-----AFORTNQOD-----NRRFINDMVKIY 587
Db 1315 VANFTLRLEQRQWKVTVNGVDMKLPVVLANGQIRASQHGSDVVIETDFGLRVAYDLV--Y 1372
QY 588 SITATNAVGDVASSCRACA-----LGSEQS-----SSCV---617
Db 1373 YVRVTVEGNYIQQMCGLCGNYNGDPKDDFOKPNGSQAGNANEFNGSWEEVVPDSPCLPPT 1432
QY 618 PCPPGHYIEKETNOCKECPDPTYLSIHQVYGKEACIPCGPGSKNNQDHSVCY-----669
Db 1433 PCPPG---SEDCIPSHKCPPE-----LEKKYQKEEF--CGLLSSPTGPLSSCHKLVDPQGP 1483
QY 670 -SDCFFYHEKENQILHYDFSNLSSVGLMNGPSFTSKGTKYFHHFFNISLCGHEGKKMALC 728
Db 1484 LKDCIF-----DLCLGGGNLSILC 1502
QY 729 TNNITDFTVKEIVAGSDDYTNLVGAFV--CQST---IIPSESKGFRAALSSQSIIILADTF 783
Db 1503 SN-----IHAYVSACQAAGGHVEPWRTETF-----1527
QY 784 IGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSATTSCINGRSTAVKMRCPNPTKSG 843
Db 1528 -----CPMECPPNSHYELCADTCSLGC-----1549
QY 844 AGVISVPSKCPAGTCDG--CTFYFLWESAEACPL-----875
Db 1550 -SALSAPQCQDGCAGCQCDSDGFLYNGQACVPIQQCGCYHNGVYYPEQTVLIDNCRQQ 1608
QY 876 -----CTEHD-----FHEI-----EGACKRG 891
```



Db 1609 CTCHAGKGMVQEHSCPKGQVCQPSGGILSCVTKDPCHGVTCRPQETCKEQGGQGVCLPN 1668  
QY 892 FQETLYVWNEPKW 904  
Db 1669 YEATCWLWGDPHY 1681

RESULT 14  
US-08-185-432-18  
; Sequence 18, Application US/08185432  
; Patent No. 5750652  
; GENERAL INFORMATION:  
; APPLICANT: Artavanis-Tsakonas, Spyridon  
; APPLICANT: Busseau, Isabelle  
; APPLICANT: Diederich, Robert J.  
; APPLICANT: Xu, Tian  
; APPLICANT: Matsuno, Kenji  
; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND  
; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/185,432  
; FILING DATE: 21-JAN-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 7326-006  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2523 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-185-432-18

Query Match 2.6%; Score 149.5; DB 1; Length 2523;  
Best Local Similarity 19.0%; Pred. No. 0.00062;  
Matches 168; Conservative 83; Mismatches 284; Indels 349; Gaps 50;

QY 278 NITIEGVAYTSEC--FPCK-PGTFENKPGSFNCQVCPNRY-----SEKGAKECI-- 324  
Db 675 NINI-----NECDSNPCRNGGTCCKDQINGFTC-VCPDGYHDMCLSEVNECNSNPCIHG 727  
QY 325 -----RCKDDSQFSGS-----SECTERP-----PCTTKDYFQ---- 351  
Db 728 ACHDGVNGYKCDCEAGWSGSNCDDINNNECESNPCMNGGTCKDMTGAYICTCKAGFSGPNC 787  
QY 352 -----IHTPCDEEGKTQIMYKWIETKICRED-----LTDAILRPPSGE 389  
Db 788 QTNINECSSNCLNHG-----TCIDDDVAGYKCNCLMPLPYTGAIACEAVLAPCAGS 835  
QY 390 KDCPPPCNPFGYNNSS-----SCHPCPPG-----TFSDGTKEC---RPCPAGTEPALGFEY 438  
Db 836 -----PCKNGGRCKESEDFTFSCC-CPPGWQGGTCEIDMNECVNRPGRNG----- 880

QY 439 KWNVLPGNMKTSCFNV-GNSKCDGMNGW-----EVAGDHQSGA---GGSDNDYLIILNL 489  
Db 881 -----ATCQNTNGSYKCNCKPGYTGRNCEMDIDDCQPNCHNGGSCSDG--INM 927  
QY 490 HI-----PGFKPPTSMTGATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVESNGGT 545  
Db 928 FFCNCPAGFRGPK-----CEE-----DINECASNPCKNGA-- 957  
QY 546 KEQAYTHIIFKNATFTTWFQRTNQGDNRFRINDMKIYSITATNAVQV--ASSCR 603  
Db 958 -----NCT-----DCVNSYTCTCQPGFSGIHCESNTP 984  
QY 604 ACALGSEQSGSSCV-----PCPP--GHYIEKETNOCKECPDPTYLSIHQVYVKEAC 652  
Db 985 DCTESSCFNGGTCIDGINTFTCQCPPGFTGSYQCHDINECDKPCNLNGGTCQDSYGYK 1044  
QY 653 I-----PCGPGSKNNQDHSVCYSDCFYHEKENQILHYDFSNLSS 692  
Db 1045 TCPQGYTGLNQNQLVRWCDSSPCKNGGKCQWQTNFYRCEC-----KSGW 1088  
QY 693 VGSLMNGPSFT-----SKGTKYFHFH-NISLC-----CHEGKKMALCTNNI 732  
Db 1089 TGVYCDVPSVSCVAAKQQGVVDIVHLCRNSGMCVDTGNTHECRQAGYTG---SYCEEQV 1145  
QY 733 TDETVKEIVAGSDDYTNLVGAFVQCSTIIPSESKGFRAALSQSIIADTFIGTVETTL 792  
Db 1146 DECSPNPCQNGA-TCTDYLGGYSCECV-----AGYHGVNCSBEEINECLSHPCQNGGTCTI 1198  
QY 793 KNINIKEDMFPVPTS-----QIPDVHFFYKSTATTSCINGRSTAVKMRNPTKSGAV 846  
Db 1199 DLINTYKCSGPRGTQGVHCEINVDCTPFYDSFTLEPKCFNNG-----KCIDRVGGYNC 1252  
QY 847 ISVPSKCPAGTCDGCTFYFLWESAFAEACPLCTEHDFFE-IEGAC-KRGEQETLYVWNEPKW 904  
Db 1253 I-----CPPG-----FVGECG-----DVNECLSNPCDSRGTCQNCIQLVNDYRC 1292  
QY 905 -CIKGIS-----LPEKKLATCETVD-----FWLKVAGAGVGAFTAVLLVALTTC 945  
Db 1293 ECRQGTGRRCESVVDGCKGMPCRNGGTCAVASNTERGEFICKPPGFDGATCE-YDSRTC 1351  
QY 946 YFWKKNOKLEYKYSKLVMTTNSK-----ECELPAADSCA 979  
Db 1352 ----SNLRCQNGGTCISVLTSSKVCVSEGYTGATCQYPVISPCA 1391

RESULT 15  
US-08-899-232-3  
; Sequence 3, Application US/08899232  
; Patent No. 6436650  
; GENERAL INFORMATION:  
; APPLICANT: Artavanis-Tsakonas, Spyridon  
; APPLICANT: Qi, Huilin  
; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON  
; FILE REFERENCE: 7326-046  
; CURRENT APPLICATION NUMBER: US/08/899,232  
; CURRENT FILING DATE: 1997-07-23  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 2523  
; TYPE: PRT  
; ORGANISM: xenopus sp.  
US-08-899-232-3

Query Match 2.6%; Score 149.5; DB 4; Length 2523;  
Best Local Similarity 19.0%; Pred. No. 0.00062;  
Matches 168; Conservative 83; Mismatches 284; Indels 349; Gaps 50;

QY 278 NITIEGVAYTSEC--FPCK-PGTFENKPGSFNCQVCPNRY-----SEKGAKECI-- 324  
Db 675 NINI-----NECDSNPCRNGGTCCKDQINGFTC-VCPDGYHDMCLSEVNECNSNPCIHG 727  
QY 325 -----RCKDDSQFSGS-----SECTERP-----PCTTKDYFQ---- 351

Db 728 ACHDGVNGYKDCCEAGWSGNSCDINNNECESNPCMNGGTCKDMTGAYICTCKAGFSGPNC 787  
QY 352 -----IHTPCDEEGKTQIMYKWIPEKICRED-----LTDAIRLPSPGE 389  
Db 788 QTNINECSSNPCLNHG-----TCIDDVAGYKNCMLPYTGAICEAVLAPCAGS 835  
QY 390 KKDCPPPCNPGFYNNSS-----SCHPCPPG-----TFSDGTKEC--RPCPAGTEPALGFY 438  
Db 836 -----PCKNGGRCKESEDFTFSCE-CPPGWQQTCEIDMNECVNRPCRN----- 880  
QY 439 KWNVLPGNMKTSCFNV-GNSKCDGMNGW-----EVAGDHIQSGA---GGSDNDYLILNL 489  
Db 881 -----ATCQNTNGSYKCNCKPGYTGRNCEMDIDDCQPNPCHNGGSCSDG--INM 927  
QY 490 HI----PGFKPPTSMTGATGSELGRITFVFETLCSADCVLVFMVDINRKSTNVVESWGGT 545  
Db 928 FECNCPAGFRGPK-----CEE-----DINECASNPCKNGA-- 957  
QY 546 KEQAYTHIEFKNATFTFWAFQRTNQODNRRFINDMVKIYSITATNAVDGV--ASSCR 603  
Db 958 -----NCT-----DCVNSYTCTCQPGFSGIHCESTP 984  
QY 604 ACALGSEQSGSSCV-----PCPP---GHYIEKETNOCKECPDPTYLSIHQVYVYKKEAC 652  
Db 985 DCTESSCFNGGTCIDGINTFTCQCPPPGFTGSYQCQHDINECDKPKCLNGGTCQDSYGTYKC 1044  
QY 653 I-----PCGPGSKNNQDHSVYSDCFFYHEKENQILHYDFSNLSS 692  
Db 1045 TCPQGYTLNCQNLVRWCDSPCKNGGKCWQTNNFYRCEC-----KSGW 1088  
QY 693 VGSMLMGPSFT-----SKGTKYFHFF-NISLC-----GHEGKKMALCTNNI 732  
Db 1089 TGVYCDVPSVCEVAAKQOGVDIVHLCRNSGMCVDTGNTHFRCQAGYTG---SYCEEQV 1145  
QY 733 TDFTVKEIVAGSDDYTNLVGAFVQCSTIIPSESKGFRAALSSQSIILADTFIGTVVETTL 792  
Db 1146 DECSNPQCQNGA-TCTDYLGGYSCECV-----AGYHGVNCSEEEINECLSHPCQNGGTCTI 1198  
QY 793 KNINIKEDMFPVPTS-----QIPDVHFFYKSSATTSCINGRSTAVKMRCNPTKSGAGV 846  
Db 1199 DLINTYKCSCPRGQTQGVHCEINVDCTPFYDSFTLEPKCFNNG-----KCIDRVGGYNC 1252  
QY 847 ISVPSKCPAGTCDGCTFYFLWESAECPLCTEHDFFE-IEGAC-KRGFQETLYVWNEPKW 904  
Db 1253 I-----CPPG-----FVGERCEG-----DVNECLSNPCDSRGTQNCIQLVNDYRC 1292  
QY 905 -CIKGIS-----LPEKKLATCETVD-----FWLKVAGVGAGTAVLLVALTC 945  
Db 1293 ECRQGTGRRCESVVDGCKGMPCRNGGTCAVASNTERGTICKCPPPGFDGATCE-YDSRTC 1351  
QY 946 YFWKKNQKLEYKYSKLVMTTNSK-----ECELPAADSCA 979  
Db 1352 ----SNLRCQNGGTCTISVLTSSKVCSEGYTGATCQYVISPSCA 1391



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using 'sw model

Run on: May 12, 2003, 13:21:38 ; Search time 188.369 Seconds  
(without alignments)  
501.729 Million cell updates/sec

Title: US-10-073-333A-4  
Perfect score: 5681  
Sequence: 1 MLFRARGPVRGWRGWRPAEA.....KEKEDHFESVQLKTSRPNI 1027

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 349150.seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5681	100.0	1027	9	US-10-140-164-4
2	5307	93.4	963	9	US-10-140-164-2
3	3025	53.2	1013	9	US-10-028-072-38
4	3025	53.2	1013	9	US-10-121-049-38
5	3025	53.2	1013	9	US-10-123-904-38
6	3025	53.2	1013	9	US-10-140-470-38
7	3025	53.2	1013	9	US-10-175-746-38
8	3025	53.2	1013	9	US-10-176-918-38
9	3025	53.2	1013	9	US-10-176-921-38
10	3025	53.2	1013	9	US-10-137-865-38
11	3025	53.2	1013	9	US-10-140-474-38
12	3025	53.2	1013	9	US-10-142-431-38
13	3025	53.2	1013	9	US-10-143-114-38
14	3025	53.2	1013	9	US-10-140-002-38
15	3025	53.2	1013	9	US-10-142-419-38
16	3025	53.2	1013	9	US-10-123-262-38
17	3025	53.2	1013	9	US-10-142-423-38
18	3025	53.2	1013	9	US-10-121-050-38
19	3025	53.2	1013	9	US-10-141-755-38

20	3025	53.2	1013	9	US-10-143-032-38	Sequence 38, Appl
21	3025	53.2	1013	9	US-10-123-108-38	Sequence 38, Appl
22	3025	53.2	1013	9	US-10-123-236-38	Sequence 38, Appl
23	3025	53.2	1013	9	US-10-123-261-38	Sequence 38, Appl
24	3025	53.2	1013	9	US-10-140-921-38	Sequence 38, Appl
25	3025	53.2	1013	9	US-10-140-928-38	Sequence 38, Appl
26	3025	53.2	1013	9	US-10-121-045-38	Sequence 38, Appl
27	3025	53.2	1013	9	US-10-123-292-38	Sequence 38, Appl
28	3025	53.2	1013	9	US-10-123-903-38	Sequence 38, Appl
29	3025	53.2	1013	9	US-10-124-819-38	Sequence 38, Appl
30	3025	53.2	1013	9	US-10-124-822-38	Sequence 38, Appl
31	3025	53.2	1013	9	US-10-140-925-38	Sequence 38, Appl
32	3025	53.2	1013	9	US-10-160-498-38	Sequence 38, Appl
33	3025	53.2	1013	9	US-10-121-041-38	Sequence 38, Appl
34	3025	53.2	1013	9	US-10-121-043-38	Sequence 38, Appl
35	3025	53.2	1013	9	US-10-121-047-38	Sequence 38, Appl
36	3025	53.2	1013	9	US-10-123-215-38	Sequence 38, Appl
37	3025	53.2	1013	9	US-10-123-902-38	Sequence 38, Appl
38	3025	53.2	1013	9	US-10-123-908-38	Sequence 38, Appl
39	3025	53.2	1013	9	US-10-123-909-38	Sequence 38, Appl
40	3025	53.2	1013	9	US-10-123-910-38	Sequence 38, Appl
41	3025	53.2	1013	9	US-10-124-813-38	Sequence 38, Appl
42	3025	53.2	1013	9	US-10-124-817-38	Sequence 38, Appl
43	3025	53.2	1013	9	US-10-124-824-38	Sequence 38, Appl
44	3025	53.2	1013	9	US-10-125-922-38	Sequence 38, Appl
45	3025	53.2	1013	9	US-10-125-924-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1  
US-10-140-164-4  
; Sequence 4, Application US/10140164  
; Publication No. US20030072736A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker et al.  
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16  
; FILE REFERENCE: PF514C1  
; CURRENT APPLICATION NUMBER: US/10/140,164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 1027  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-140-164-4

Query Match 100.0%; Score 5681; DB 9; Length 1027;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1027; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFRARGPVRGWRGWRPAEAPRRGRSPWPSPAWICCWALAGCQAAWAGDLPSSSSRPLPP 60  
|||||  
Db 1 MLFRARGPVRGWRGWRPAEAPRRGRSPWPSPAWICCWALAGCQAAWAGDLPSSSSRPLPP 60



QY 61 QCEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNOVC 120  
Db 61 QCEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNOVC 120  
QY 121 SKCGEGTYSLSGSGIKFDEWDELPAAGFNSNIATFMDTVVGPSPDSRPPGCGNNSSWIPRGNVIE 180  
Db 121 SKCGEGTYSLSGSGIKFDEWDELPAAGFNSNIATFMDTVVGPSPDSRPPGCGNNSSWIPRGNVIE 180  
QY 181 SNRDDCTVSLIYAVHLKSGYVFFFEYQYVDNNIFFEFFFIONDQCQEMDTTDDKWKVLTND 240  
Db 181 SNRDDCTVSLIYAVHLKSGYVFFFEYQYVDNNIFFEFFFIONDQCQEMDTTDDKWKVLTND 240  
QY 241 GEWGSVLMKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECPCPKPGTFSN 300  
Db 241 GEWGSVLMKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECPCPKPGTFSN 300  
QY 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360  
Db 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360  
QY 361 KTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPCPNPGFYNNNGSSSCHPCPPGTFSDGT 420  
Db 361 KTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPCPNPGFYNNNGSSSCHPCPPGTFSDGT 420  
QY 421 KECRPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480  
Db 421 KECRPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480  
QY 481 DNDYLILNLHIPGFKPPTSMGTGATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540  
Db 481 DNDYLILNLHIPGFKPPTSMGTGATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540  
QY 541 SWGGTKEKQAYTHIIFKNATFTFTWAFORTNOGQDNRFRFINDMVKIYSITATNAVDGVAS 600  
Db 541 SWGGTKEKQAYTHIIFKNATFTFTWAFORTNOGQDNRFRFINDMVKIYSITATNAVDGVAS 600  
QY 601 SCRACALGSEQSGSSCPCPPGHYIEKETNQCKECPDPTYLSIHQVYGKEACIPCGPGSK 660  
Db 601 SCRACALGSEQSGSSCPCPPGHYIEKETNQCKECPDPTYLSIHQVYGKEACIPCGPGSK 660  
QY 661 NNQDHSVVCYSDCFYHEKENQILHYDEFNLSSVGLMNGPSFTSKGTKYFHHFNISLCGH 720  
Db 661 NNQDHSVVCYSDCFYHEKENQILHYDEFNLSSVGLMNGPSFTSKGTKYFHHFNISLCGH 720  
QY 721 EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFVQCSTIIPSESKGFRAALSSQSIILA 780  
Db 721 EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFVQCSTIIPSESKGFRAALSSQSIILA 780  
QY 781 DTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSATTSCINGRSTAVKMRNPT 840  
Db 781 DTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSATTSCINGRSTAVKMRNPT 840  
QY 841 KSGAGVISVPSKCPAGTCDCGTFFYFLWESAECPLCTEHDHFEIEGACKRGFOETLYVWN 900  
Db 841 KSGAGVISVPSKCPAGTCDCGTFFYFLWESAECPLCTEHDHFEIEGACKRGFOETLYVWN 900  
QY 901 EPKWKIGISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKNQKLEYKYSK 960  
Db 901 EPKWKIGISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKNQKLEYKYSK 960  
QY 961 LVMTNSKECELPAADSCAIMEGEDNEEVVYSNKQSLGKLSLATKEKEDHFESVOLK 1020  
Db 961 LVMTNSKECELPAADSCAIMEGEDNEEVVYSNKQSLGKLSLATKEKEDHFESVOLK 1020  
QY 1021 TSRSNPNI 1027  
Db 1021 TSRSNPNI 1027

RESULT 2  
US-10-140-164-2  
; Sequence 2, Application US/10140164  
; Publication No. US20030072736A1

; GENERAL INFORMATION:  
; APPLICANT: Baker et al;  
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor Trl16  
; FILE REFERENCE: PF514C1  
; CURRENT APPLICATION NUMBER: US/10/140,164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 963  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-140-164-2

Query Match 93.4%; Score 5307; DB 9; Length 963;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 953; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFRARGPVGRGWGRPAEAPRRGRSPWPSPAWICCWALAGCAAAGDLPSSSSRPLPP 60  
Db 1 MLFRARGPVGRGWGRPAEAPRRGRSPWPSPAWICCWALAGCAAAGDLPSSSSRPLPP 60  
QY 61 QCEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNOVC 120  
Db 61 QCEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNOVC 120  
QY 121 SKCGEGTYSLSGSGIKFDEWDELPAAGFNSNIATFMDTVVGPSPDSRPPGCGNNSSWIPRGNVIE 180  
Db 121 SKCGEGTYSLSGSGIKFDEWDELPAAGFNSNIATFMDTVVGPSPDSRPPGCGNNSSWIPRGNVIE 180  
QY 181 SNRDDCTVSLIYAVHLKSGYVFFFEYQYVDNNIFFEFFFIONDQCQEMDTTDDKWKVLTND 240  
Db 181 SNRDDCTVSLIYAVHLKSGYVFFFEYQYVDNNIFFEFFFIONDQCQEMDTTDDKWKVLTND 240  
QY 241 GEWGSVLMKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECPCPKPGTFSN 300  
Db 241 GEWGSVLMKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECPCPKPGTFSN 300  
QY 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360  
Db 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360  
QY 361 KTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPCPNPGFYNNNGSSSCHPCPPGTFSDGT 420  
Db 361 KTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPCPNPGFYNNNGSSSCHPCPPGTFSDGT 420  
QY 421 KECRPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480  
Db 421 KECRPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480  
QY 481 DNDYLILNLHIPGFKPPTSMGTGATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540  
Db 481 DNDYLILNLHIPGFKPPTSMGTGATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540  
QY 541 SWGGTKEKQAYTHIIFKNATFTFTWAFORTNOGQDNRFRFINDMVKIYSITATNAVDGVAS 600  
Db 541 SWGGTKEKQAYTHIIFKNATFTFTWAFORTNOGQDNRFRFINDMVKIYSITATNAVDGVAS 600

\*

QY 601 SCRAQALGSEQSGSCVPCPPGHYIEKETNCKECPDPTLYLSIHQVYKEACIPCGPGSK 660  
Db 601 SCRAQALGSEQSGSCVPCPPGHYIEKETNCKECPDPTLYLSIHQVYKEACIPCGPGSK 660  
QY 661 NNQDHSVSCYDCFFYHEKENQILHYDFSLSVGSMLMNGPSTSGTKYFHFNNISLCGH 720  
Db 661 NNQDHSVSCYDCFFYHEKENQILHYDFSLSVGSMLMNGPSTSGTKYFHFNNISLCGH 720  
QY 721 EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFVCQSTIIPSESKGFRAALSSQSIILA 780  
Db 721 EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFVCQSTIIPSESKGFRAALSSQSIILA 780  
QY 781 DTFIGTVTETLKNINIKEDMFPVPPTSQIPDVHFFYKSSTATTSINGRSTAVKMRNPT 840  
Db 781 DTFIGTVTETLKNINIKEDMFPVPPTSQIPDVHFFYKSSTATTSINGRSTAVKMRNPT 840  
QY 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEHDFHEIEGACKRGFQETLYVWN 900  
Db 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEHDFHEIEGACKRGFQETLYVWN 900  
QY 901 EPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAFNAVLLVALTCTYFWKKNQK 953  
Db 901 EPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAFNAVLLVALTCTYFWKKNQK 953

RESULT 3

US-10-028-072-38  
; Sequence 38, Application US/10028072  
; Publication No. US20030004311A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang  
; TITLE OF INVENTION:  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/028,072  
; CURRENT FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/049911  
; PRIOR FILING DATE: 1997-06-18  
; PRIOR APPLICATION NUMBER: 60/056974  
; PRIOR FILING DATE: 1997-08-26  
; PRIOR APPLICATION NUMBER: 60/059113  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059115  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059117  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059122  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059184  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059352  
; PRIOR FILING DATE: 1997-09-19  
; PRIOR APPLICATION NUMBER: 60/059588  
; PRIOR FILING DATE: 1997-09-19  
; PRIOR APPLICATION NUMBER: 60/059836

; PRIOR FILING DATE: 1997-09-24  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/062285  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/062287  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/062814  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/062816  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063045  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063082  
; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: 60/063127  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063327  
; PRIOR FILING DATE: 1997-10-27  
; PRIOR APPLICATION NUMBER: 60/063329  
; PRIOR FILING DATE: 1997-10-27  
; PRIOR APPLICATION NUMBER: 60/063550  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063561  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063704  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/063733  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/063735  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/063738  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/063755  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/064248  
; PRIOR FILING DATE: 1997-11-03  
; PRIOR APPLICATION NUMBER: 60/064809  
; PRIOR FILING DATE: 1997-11-07  
; PRIOR APPLICATION NUMBER: 60/065186  
; PRIOR FILING DATE: 1997-11-12  
; PRIOR APPLICATION NUMBER: 60/065846  
; PRIOR FILING DATE: 1997-11-17  
; PRIOR APPLICATION NUMBER: 60/066364  
; PRIOR FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: 60/066453  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/066511  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/066770  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/069212  
; PRIOR FILING DATE: 1997-12-11  
; PRIOR APPLICATION NUMBER: 60/069278  
; PRIOR FILING DATE: 1997-12-11  
; PRIOR APPLICATION NUMBER: 60/069334  
; PRIOR FILING DATE: 1997-12-11  
; PRIOR APPLICATION NUMBER: 60/069694  
; PRIOR FILING DATE: 1997-12-16  
; PRIOR APPLICATION NUMBER: 60/072320  
; PRIOR FILING DATE: 1998-01-23  
; PRIOR APPLICATION NUMBER: 60/073612  
; PRIOR FILING DATE: 1998-02-04  
; PRIOR APPLICATION NUMBER: 60/074086  
; PRIOR FILING DATE: 1998-02-09  
; PRIOR APPLICATION NUMBER: 60/074092  
; PRIOR FILING DATE: 1998-02-09  
; PRIOR APPLICATION NUMBER: 60/077791  
; PRIOR FILING DATE: 1998-03-12  
; PRIOR APPLICATION NUMBER: 60/078910  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/079294  
; PRIOR FILING DATE: 1998-03-25

;	PRIOR APPLICATION NUMBER: 60/079663
;	PRIOR FILING DATE: 1998-02-27
;	PRIOR APPLICATION NUMBER: 60/079728
;	PRIOR FILING DATE: 1998-03-27
;	PRIOR APPLICATION NUMBER: 60/080165
;	PRIOR FILING DATE: 1998-03-31
;	PRIOR APPLICATION NUMBER: 60/081203
;	PRIOR FILING DATE: 1998-04-09
;	PRIOR APPLICATION NUMBER: 60/081229
;	PRIOR FILING DATE: 1998-04-09
;	PRIOR APPLICATION NUMBER: 60/081695
;	PRIOR FILING DATE: 1998-04-14
;	PRIOR APPLICATION NUMBER: 60/081817
;	PRIOR FILING DATE: 1998-04-15
;	PRIOR APPLICATION NUMBER: 60/081818
;	PRIOR FILING DATE: 1998-04-15
;	PRIOR APPLICATION NUMBER: 60/082999
;	PRIOR FILING DATE: 1998-04-24
;	PRIOR APPLICATION NUMBER: 60/083322
;	PRIOR FILING DATE: 1998-04-28
;	PRIOR APPLICATION NUMBER: 60/0833545
;	PRIOR FILING DATE: 1998-04-29
;	PRIOR APPLICATION NUMBER: 60/084600
;	PRIOR FILING DATE: 1998-05-07
;	PRIOR APPLICATION NUMBER: 60/084627
;	PRIOR FILING DATE: 1998-05-07
;	PRIOR APPLICATION NUMBER: 60/084637
;	PRIOR FILING DATE: 1998-05-07
;	PRIOR APPLICATION NUMBER: 60/085149
;	PRIOR FILING DATE: 1998-05-12
;	PRIOR APPLICATION NUMBER: 60/085323
;	PRIOR FILING DATE: 1998-05-13
;	PRIOR APPLICATION NUMBER: 60/085338
;	PRIOR FILING DATE: 1998-05-13
;	PRIOR APPLICATION NUMBER: 60/085339
;	PRIOR FILING DATE: 1998-05-13
;	PRIOR APPLICATION NUMBER: 60/085579
;	PRIOR FILING DATE: 1998-05-15
;	PRIOR APPLICATION NUMBER: 60/085697
;	PRIOR FILING DATE: 1998-05-15
;	PRIOR APPLICATION NUMBER: 60/085704
;	PRIOR FILING DATE: 1998-05-15
;	PRIOR APPLICATION NUMBER: 60/086414
;	PRIOR FILING DATE: 1998-05-22
;	PRIOR APPLICATION NUMBER: 60/086430
;	PRIOR FILING DATE: 1998-05-22
;	PRIOR APPLICATION NUMBER: 60/087106
;	PRIOR FILING DATE: 1998-05-28
;	PRIOR APPLICATION NUMBER: 60/088026
;	PRIOR FILING DATE: 1998-06-04
;	PRIOR APPLICATION NUMBER: 60/088730
;	PRIOR FILING DATE: 1998-06-10
;	PRIOR APPLICATION NUMBER: 60/088741
;	PRIOR FILING DATE: 1998-06-10
;	PRIOR APPLICATION NUMBER: 60/088810
;	PRIOR FILING DATE: 1998-06-10
;	PRIOR APPLICATION NUMBER: 60/088858
;	PRIOR FILING DATE: 1998-06-11
;	PRIOR APPLICATION NUMBER: 60/089532
;	PRIOR FILING DATE: 1998-06-17
;	PRIOR APPLICATION NUMBER: 60/089599
;	PRIOR FILING DATE: 1998-06-17
;	PRIOR APPLICATION NUMBER: 60/089907
;	PRIOR FILING DATE: 1998-06-18
;	PRIOR APPLICATION NUMBER: 60/089947
;	PRIOR FILING DATE: 1998-06-19
;	PRIOR APPLICATION NUMBER: 60/090349
;	PRIOR FILING DATE: 1998-06-23
;	PRIOR APPLICATION NUMBER: 60/090429
;	PRIOR FILING DATE: 1998-06-24
;	PRIOR APPLICATION NUMBER: 60/090445
;	PRIOR FILING DATE: 1998-06-24
;	PRIOR APPLICATION NUMBER: 60/090538

;  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090863  
;  
; PRIOR FILING DATE: 1998-06-26  
; PRIOR APPLICATION NUMBER: 60/091360  
;  
; PRIOR FILING DATE: 1998-07-01  
; PRIOR APPLICATION NUMBER: 60/091519  
;  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091982  
;  
; PRIOR FILING DATE: 1998-07-07

Query Match	53.2%;	Score 3025;	DB 9;	Length 1013;
Best Local Similarity	54.4%;	Pred. No. 1.3e-228;		
Matches 535;	Conservative 169;	Mismatches 265;	Indels 14	

QY	46	WAG--DLPSSSRPLPPCQBKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKE	102
Db	31	WAGTAFQVTOGTGPELHACKSEYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKGT	90
QY	103	CTFSCASGEYLEMKNOVCSCGEGTYSLGSGIKFDEWDELPAFGFSNIATFMDTVVGP	162
Db	91	CSFSCNAGEFLDMKDQSCKPCAEGRYSLGTGIRFDEWDELPIGCFASLSANMELDD	150
QY	163	RPDGCNNSWIPRGNYIESNRDDCTVSLIYAVHLKKSGYVFEYQYVDNNIEFEFFIQ	222
Db	151	-TGNCSTSSKWVPRGDYIASNTDECTATLMYAVNLKQSGTVNFEEYYPDSSIIEFF	209
QY	223	QCQEMDTTDTKWVKLTDNGEWGSHSVMLKSGTNILYWRTTGILMGSKAVKPV	282
Db	210	QCQP-NADDSRWMTTEKG-WEPHSELNRGNVLYWRTTAFSVMTKVPKPVLRNIAIT	267
QY	283	GVAYTSECFFPKPGTFSNKPGSFNQVCPRNTYSEKGAKECIRCKDDSQFS--GSSE	340
Db	268	GVAYTSECFFPKPGTYADKQGSSEFCKLCPANSYSNKGETSCHQC-DPDKYSEKGS	326
QY	341	RPPCTTKDYFIHTPCDEEGKTOIMYKWIPEKICREDLTDAIRLPSPGSEKKDCPP	400
Db	327	RPACTDKDYFYTHTACDANGETQLMYKWKAPKICSEDLGAVKLPASGVKTHCPP	386
QY	401	YNNGSSSCHPCPPGTFSDGTKECRPCPACTEPALGFYKWNVLPGNMKTSFCNVG	460
Db	387	EKTNNSTCQPCPYGSNGS-DCTRCPACTEPAGVFYKWNLTPTNMETVLSGINFEY	445
QY	461	DGMNGEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTS-MTGATGSELGRITFVF	519
Db	446	KGMTGWEVAGDHIYTAAGASDNDFMILTLLVVGFRPPOSVMADTENKEVARITF	505
QY	520	SADCVLYEMVDINRKSTNVVESWGCTKEKQAYTHIIFKNATFTFTWAFORTNQO	579
Db	506	SVNCELYFMVGNSRNTNPVETWKSGKGOSYTYIIIEENTTTSTWAFORTTFHEA	565
QY	580	INDMVKIYSTATNAVDGVASSCRACALGSEQSGSSCVPCPPGHYIEKETNQCKE	639
Db	566	TNDVAKIYSINVTNVMNGVASYCRPCALEASDVSSCTSCPAGYYIDRDSGTCHS	625
QY	640	YLSIHQVYGKEACIPCGPGSKNNQDHSVCYSDGFFYHEKENQILHYDFSNLSSV	699
Db	626	ILKAHQPYGVQACVPCGPGTKNKHLSLCYNDCTESRNTPTRTFNYSFALANTVT	685
QY	700	PSFTSKGTYKHYEFENISLCGHEGKKMALCTNNTDTFTVKEIVAGSDDYTNLGA	759
Db	686	PSFTSKGLKYFHUFTLSLCGNQGRKMSVCTDNVTDLRIPE--GESGFKSITAYV	742
QY	760	IIPSESKGFRAALSSQSILLADTFIGVTVETTLKNINIKEDMFPVPPTSQIPDV	819
Db	743	IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSAPAEFLHLESIGIPDV	802
QY	820	TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESAECPL	879
Db	803	DVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCDGTCDGCNFHFLWESAAAC	862
QY	880	DFHEIEGACKRGFOETLYVWNEPKWCIKGISLPEKKLATCETVDFWLKVAGVG	939
Db	863	DVHAIVSSCVAGIOXTYVXREPKLCSGGISLPEORTYICKTIDFWLKVGISAGT	922



QY 940 LVALTCYFWKKNOKLEYKYSKLVMTTNSKECELPAAADSCAIMEGEDNEEEVVSNNKQSL 999  
Db 923 LTVLTCTYFWKKNOKLEYKYSKLVMTNATLKDCDLPAAADSCAIMEGEDVEDDLIFTSKSLF 982  
QY 1000 GKLKSLATKEKEDHFESVQLKTS 1022  
Db 983 GKIKSFTSKRTPDGFDSVPLKTS 1005  
RESULT 4  
US-10-121-049-38  
; Sequence 38, Application US/10121049  
; Publication No. US2003002239A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3330R1C17  
; CURRENT APPLICATION NUMBER: US/10/121,049  
; CURRENT FILING DATE: 2002-04-12  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 38  
; LENGTH: 1013  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 877, 882  
; OTHER INFORMATION: unknown amino acid  
US-10-121-049-38

Query Match 53.2%; Score 3025; DB 9; Length 1013;  
Best Local Similarity 54.4%; Pred. No. 1.3e-228;  
Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;  
QY 46 WAG---DLPSSSRPLPPCQEKDYHEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKE 102  
Db 31 WAGTAFQVTQGTGPELHACKSEYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKGT 90  
QY 103 CTFSCASGEYLEMKNQVCSKCGEGTYSLSGSGIKFDEWDELPAAGFSNIATFMDTVVGP 162  
Db 91 CSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSA 150  
QY 163 RPDGNNSSWIPRGNIESNRDDCTVSLIYAVHLKSGYVFEYQYVDNNIFFEFFIQND 222  
Db 151 -TGNCSTSKWVPRGDYIASNTDECTATLMYAVNLKQSGTVNFEYYPDSSIIEFFVQND- 209  
QY 223 QCQEMDTTDDKWVLTNDGEWGSVMLKSGTNILYWRRTTGILMSKAVKPKVLKNITIE 282  
Db 210 QCQP-NADDSRWKMTTEKG-WEFHSELNRGNVLYWRRTTAFSVWTKVPKPLVRNIAIT 267  
QY 283 GVAYTSECFPCPKPGTFSNKPQSFNCQVCPRNTYSEKGAKEICRCKDDSQFS--GSSECTE 340  
Db 268 GVAYTSECFPCPKPGTYADKQKQSSFCCKLCPANSYSNKGTSCHQC-DPDKYSEKGS 326

QY 341 RPPCTTKDYFQIHTPCDEEGKQIIMYKWKIEPKICREDLTDAIRLPPSGEKKDCPPCPNPGF 400  
Db 327 RPACTDKDYFYTHTACDANGETQLMYKWKAPKTCSEDLGAVKLPASGVKTHCPPCNPGF 386  
QY 401 YNNGSSSCHPCPPGTFSDGTEKCRPCPCAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKC 460  
Db 387 EKTNNSTCQPCPGYSYNGS-DCTRCPCAGTEPAVGFEYKWNVLPNTMETTVLSGINFEY 445  
QY 461 DGMNGWEVAGDHIOGAGGSDNDYLIILNLHLPDGPKPPPTS-MTGATGSELGRITFFVETLC 519  
Db 446 KGMTGWEVAGDHIYTAAGASDNDFMILTLPVPGFRPPQSVMDATENKEVARITFFVETLC 505  
QY 520 SADCVLYFMVDINRKSTNVVSWGGTKEKQAYTHIIFKNATFTFTWAFORTNQGDNRRE 579  
Db 506 SVNCELYFMVGVNSRTNTPVETWKGSKGQSYTYIEENTTTSTWAFORTTFHEASRKY 565  
QY 580 INDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKETNQCKECPDPT 639  
Db 566 TNDVAKIYSINVTNVMNGVASYCRPCALEASDVGSSTSCPAGYIIDRSGTCHSCPNT 625  
QY 640 YLSIHQVYGKEACIPCGPGSKNNQDHSVCYSDCFFYHEKENQILHYDFSLSVSGSLMNG 699  
Db 626 ILKAHQPYGVQACVPCGPGTKNNKIHSCLYNDCTFSRNTPTRTFNYSALANTVTLGG 685  
QY 700 PSETSKGTKYFHFNFNISLCHGHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFVCQST 759  
Db 686 PSFTSKGLKYFHHFTLSLCGNQGRKMSVCTDNVTDLRIPE--GESGFSKSIYAYVCQAV 742  
QY 760 IIPSESKGFRAALSSQSIILADTFIGTVVETTLKNINIKEDMFPVPTSQIPDVHFFYKSS 819  
Db 743 IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSAPAELEHLESLGIPDVIFFYRSN 802  
QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEH 879  
Db 803 DVTQSCSSGRSTIRVRCSPQKTVPGSLLPLPGTCDGTCSDGTCDFHFLWESAAACPLCSVA 862  
QY 880 DFHEIEGACKRGFQETLYVWNEPKWCIKIGISLPEKKLATCETVDFWLKVGAGVGAFTAVL 939  
Db 863 DYHAIIVSSCVAGIQXTTYVXREPKLCSGGTSLPEQRTVICKTIDFWLKVGISAGTCTAIL 922  
QY 940 LVALTCYFWKKNOKLEYKYSKLVMTTNSKECELPAAADSCAIMEGEDNEEEVVSNNKQSL 999  
Db 923 LTVLTCTYFWKKNOKLEYKYSKLVMTNATLKDCDLPAAADSCAIMEGEDVEDDLIFTSKSLF 982  
QY 1000 GKLKSLATKEKEDHFESVQLKTS 1022  
Db 983 GKIKSFTSKRTPDGFDSVPLKTS 1005

RESULT 5  
US-10-123-904-38  
; Sequence 38, Application US/10123904  
; Publication No. US20030022328A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME







Db 743 IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSPAELFHLESGLGIPDVIFFYRSN 802  
Qy 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPKCPAGTCDCGTFFYFLWESAECPLCTEH 879  
Db 803 DVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCSDGTCDCGNFHLWESAAACPLCSVA 862  
Qy 880 DFHEIEGACKRGFQETLYVWNEPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAFTAVL 939  
Db 863 DYHAIVSSCVAGIQXTTYVXREPKLCSGGISLPEQORVTICKTIDFWLKVGISAGTCTAIL 922  
Qy 940 LVALTCYFWKKNQKLEYKYSKLVMNTNSKECELPAADSCAIMEGEDNEEEVVYSNKQSLL 999  
Db 923 LTVLTCYFWKKNQKLEYKYSKLVMNATLKDCDLPAADSCAIMEGEDVEDDLIFTSKKSLE 982  
Qy 1000 GKLSLATKEKEDHFESVOLKTS 1022  
Db 983 GKIKFTSKRTPDGFDSVPLKTS 1005

RESULT 8

US-10-176-918-38  
; Sequence 38, Application US/10176918  
; Publication NO. US20030027275A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3330R1C382  
; CURRENT APPLICATION NUMBER: US/10/176,918  
; CURRENT FILING DATE: 2002-06-20  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 38  
; LENGTH: 1013  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 877, 882  
; OTHER INFORMATION: unknown amino acid  
US-10-176-918-38

Query Match 53.2%; Score 3025; DB 9; Length 1013;  
Best Local Similarity 54.4%; Pred. No. 1.3e-228;  
Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;

Qy 46 WAG---DLPSSSRPLPPCQEKDYHFYETCEDSSGSRWRVAIPNSAVDCSGLPDPVRGKE 102  
Db 31 WAGTAFQVTOGTGPELHACKESEYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKGTE 90  
Qy 103 CTFSCASGEYLEMKNQVCSKCGEGTSLGSGIKFDEWDELPAAGESNIATFMDTVVGPSPDS 162  
Db 91 CSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAES 150  
Qy 163 RPDGCNNSWIPRGNYYIESNRDDCTVSLIYAVHLKKSGYVFFEEYQYVDNNIFFEFFFIOND 222

Db 151 -TGNCITSSKWVPRGDYIASNTDECTATLMYAVNLKQSGTVNFEYYPDSSIIEFFEVQND 209  
Qy 223 QCQEMDTTDDKWKLTNDNGEWGSHSVMLKSGTNILYWRTTGILMGSKAVKPVLVKNITIE 282  
Db 210 QCQP-NADDSRWMTTEKG-WEFHSVELNRGNVLYWRTTAFSVWTKVPKPVLVNRNIAIT 267  
Qy 283 GVAYTSECFPCPKPGTFSNKPQCFNQCVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTE 340  
Db 268 GVAYTSECFPCPKPGTYADKQGSFCKLCPANSYSNKGETSCHQC-DPDKYSEKGSSSCNV 326  
Qy 341 RPPCTTKDYFQIHTPCDEEGKQIMYKWIETPKICREDLTDALRPPSGEKKDCPPCPNPGF 400  
Db 327 RPACTDKDYFYHTACDANGETQLMYKWKAPKICSEDELEGAVKLPAASGVKTHCPPCPNPGF 386  
Qy 401 YNNGSSSCHPPPGTFSDGTKECRCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKC 460  
Db 387 FKTNNSTCQPCPYGYSNGS-DCYRCPAGTEPAVGFEYKWNVLPTNNMETTVLSGINFEY 445  
Qy 461 DGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTS-MTGATGSELGRITFVFETLC 519  
Db 446 KGMTGWEVAGDHIYTAAGASDNDDEMILTLLVVPGRPPQSVMAADTENKEVARITFVFETLC 505  
Qy 520 SADCVLYFMVDINRKSTNVVSWGSGTKEKQAYTHIIFKNATFTTWAQFQRTNQGDNRRE 579  
Db 506 SVNCELYFMVGVSNTPTVETWKGSKGQSYTYIIIEENTTTTSFTWAFQRTTFHEASRKY 565  
Qy 580 INDMVKIYSITATNAVGVASSCRACALGSEQSGSCVPCPPGHYIEKETNOCKECPDPT 639  
Db 566 TNDVAKIYSINVTNVMNGVASYCRCALEASDVSSCTSCPAGYYIDRDSGTCHSCPNT 625  
Qy 640 YLSIHQVYGKEACIPCGPGSKNNDHISVCYSDCFYHEKENQILHYDFSNLSSVGLMNG 699  
Db 626 ILKAHQPGYGVQACVPCGPGTKNNKIHSCLYNDCTFSRNTPTRTFNYSALANTVTLAGG 685  
Qy 700 PSFTSKGTKYFFHFNISLCGHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFCQST 759  
Db 686 PSFTSKGLKYFHHFTLSLCGNQGRKMSVCTDNVTDLRIPE--GESGFSKSITAYVCQAV 742  
Qy 760 IIPSESKGFRAALSSQSIILADTFICVTVETTLKNINIKEDMEFPVPTSOIPDVHFFYKSS 819  
Db 743 IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSPAELFHLESGLGIPDVIFFYRSN 802  
Qy 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPKCPAGTCDCGTFFYFLWESAECPLCTEH 879  
Db 803 DVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCSDGTCDCGNFHLWESAAACPLCSVA 862  
Qy 880 DFHEIEGACKRGFQETLYVWNEPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAFTAVL 939  
Db 863 DYHAIVSSCVAGIQXTTYVXREPKLCSGGISLPEQORVTICKTIDFWLKVGISAGTCTAIL 922  
Qy 940 LVALTCYFWKKNQKLEYKYSKLVMNTNSKECELPAADSCAIMEGEDNEEEVVYSNKQSLL 999  
Db 923 LTVLTCYFWKKNQKLEYKYSKLVMNATLKDCDLPAADSCAIMEGEDVEDDLIFTSKKSLE 982  
Qy 1000 GKLSLATKEKEDHFESVOLKTS 1022  
Db 983 GKIKFTSKRTPDGFDSVPLKTS 1005

RESULT 9

US-10-176-921-38  
; Sequence 38, Application US/10176921  
; Publication NO. US20030027276A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.



```

; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C288
; CURRENT APPLICATION NUMBER: US/10/176,921
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 38
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 877, 882
; OTHER INFORMATION: unknown' amino acid
US-10-176-921-38

Query Match          53.2%; Score 3025; DB 9; Length 1013;
Best Local Similarity 54.4%; Pred. No. 1.3e-228;
Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;

QY 46 WAG---DLPSSSSRPLPPCQEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKE 102
Db 31 WAGTAFQVTOGTGPELHACKESEYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKGTE 90

QY 103 CTFSCASGEYLEMKNQVCSKCGEGTYSLSGSGTKFDEWDELPAFGFSNIATFMDTVVGPSSD 162
Db 91 CSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAAES 150

QY 163 RPDGCNNSSWIPRGNYIESNRDDCTVSLIYAVHLKKSGYVFFEYQYVDNNIFFEFFIQND 222
Db 151 -TGNCTSSKWWPRGDYIASNTDECTATLMYAVNLKQSGTVNFEYYYPDSSIIEFFVQND 209

QY 223 QCQEMDTTTTKWVKLTDNGEGSHSVMLKSGTNILYWRTTGILMGSKAVKPVLVKNITIE 282
Db 210 QCQP-NADDSRWKMTTEKG-WEFHVSVELNRGNVLYWRTTAFSVWTKVPKPVLVNRNIAIT 267

QY 283 GVAYTSECPCPKPGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFS--GSSECTE 340
Db 268 GVAYTSECPCPKPGTYADKQGSFCKLCPANSYSNKGGETSCHQC-DPDKYSEKSSSCNV 326

QY 341 RPPCTTKDYFQIHTPCDEEGKTOIMYKWIETPKICREDLTDAILRPPSGEKKDCPPCPNGF 400
Db 327 RPACTDKDYFHTACDANGETQLMYKWKAPKICSEDLEGAVKLPASGVKTHCPCPNPGF 386

QY 401 YNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFEYKWWNVLPGNMKTSCFNVGNSKC 460
Db 387 FKTNNSTCQPCPYGSYNGS-DCTRCPAGTEPAGVFEYKWWNTLPTNMETTVLSGINFEY 445

QY 461 DGMNGWEVAGDHIOGAGGSDNDYLILNLHIPGFKPPTS-MTGATSELGRITFVFETLC 519
Db 446 KGMTGWEVAGDHIYTAAGASDNDFMILTLLVPGFRPPQSVMADTENKENVARITFVFETLC 505

QY 520 SADCVLYFMVDINRKSTNVVSWGCTKEKOAYTHIIFKNATFTFTWAFQRTNQGDNRRE 579
Db 506 SVNCELYFMVGVNSRNTPTVETWKGSKQSQSYTIIENTTTSFTWAFQRTTFHEASRKY 565

QY 580 INDMVKIYSTATNAVDBGVASSCRACALGSEQSGSCVCPGPPGHYIEKETNQCKECPDPT 639
Db 566 TNDVAKIYSINVTNMVNGVASCYRCPCALEASDVGSSCTSCPAGYIIDRDSGTCHSCPNT 625

QY 640 YLSIHQVYGKEACIPCGPGSKNNQDHSVCYSDCEFFYHEKENQILHYDFSNLSSVGLMNG 699
Db 626 ILKAHQPYGVQACVPCGPGTKNNKIHSCLYNDCTFSRNTPTRTFNFSALANTVTLAGG 685
```

```

QY 700 PSFTSKGTYFHFFNISLCGHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFVCQST 759
Db 686 PSFTSKGLKYFHHFTLSLCGNQGRKMSVCTDNVTDLRIPE---GESGFSKSIYAVVCQAV 742

QY 760 IIPSESKGFRAALSSQSIILADTFIGTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSS 819
Db 743 IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGTTSPAELFHLESIGIPDVIFFYRSN 802

QY 820 TATTSCINGRSTAVKMRNMPTKSGAGVISVPSKCPAGTCDGCTFYFLWESAECPLCTEH 879
Db 803 DVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCSDGTCDCGNFHLWESAAACPLCSVA 862

QY 880 DFHEIEGACKRGFQETLYVWNEPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAFATVL 939
Db 863 DYHAIVSSCVAGIQXTYTVVXREPKLCSGGISLPEQRVTICKTIDFWLKVGISAGTCTAIL 922

QY 940 LVALTCYFWKKNQKLEYKYSKLVMTTNSKECELPAAADSCAIMEGEDNEEVVYSNKKQSL 999
Db 923 LTVLTCYFWKKNQKLEYKYSKLVMNATLKDCLPAADSCAIMEGEDVEDDLIFTSKKSLE 982

QY 1000 GKLKSLATKEKEDHFESVOLKTS 1022
Db 983 GKIKSFTSKRTPDGFDSVPLKTS 1005
```

```

RESULT 10
US-10-137-865-38
; Sequence 38, Application US/10137865
; Publication No. US20030032155A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C154
; CURRENT APPLICATION NUMBER: US/10/137,865
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 38
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 877, 882
; OTHER INFORMATION: unknown amino acid
US-10-137-865-38
```

```

Query Match          53.2%; Score 3025; DB 9; Length 1013;
Best Local Similarity 54.4%; Pred. No. 1.3e-228;
Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;

QY 46 WAG---DLPSSSSRPLPPCQEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKE 102
Db 31 WAGTAFQVTOGTGPELHACKESEYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKGTE 90

QY 103 CTFSCASGEYLEMKNQVCSKCGEGTYSLSGSGTKFDEWDELPAFGFSNIATFMDTVVGPSSD 162
```



```
Db 91 CSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAES 150
QY 163 RPDGCNNSSWIPRGNYIESNRDDCTVSLIYAVHLKKSQYVFFEYQYVDNNIFFEFFFIOND 222
Db 151 -TGNC TSSKWVPRGDYIASNTDECTATLMYAVNLKQSGTVNFEYYPDSSIIFFEFVQND 209
QY 223 QCQEMD TTTDKWVLTONGEWSHVMKSGTNILYWR TTTGILMGSKAVKPVLVKNITIE 282
Db 210 QCQP -NADDSRWMTTEKG -WEFHSVELNRGNVLYWRTTAFSVWTKVPKPVLRNIAIT 267
QY 283 GVAYTSECFPCPGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFS --GSSECTE 340
Db 268 GVAYTSECFPCPGTYADKQGSSEFCKLCPANSYSNKGGETSCHQC -DPDKYSEKSSSCNV 326
QY 341 RPPCTTKDYFQIHTPCDEEGKTQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPGE 400
Db 327 RPACTDKDYFYTHTACDANGETQLMYKWKAPKICSEDELEGAVKLPASGVKTHCPCPNPGE 386
QY 401 YNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFYKWWNVLPGNMKTSCFNVGNSKC 460
Db 387 FKTNNSTCQPCPYGSYNGS -DCTRCPAGTEPAVGFEYKWWNTLPTNMETTVLSGINFEY 445
QY 461 DGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGKPPTS -MTGATGSELGRITVFETLC 519
Db 446 KGMTGWEVAGDHIYTAAGASDNDFMILTLLVVPGRPPQSVMDATENKEVARITVFETLC 505
QY 520 SADCVLVFMVDINRKSTNVVSWGGTKEKQAYTHIIFKNATFTFTWAFQRTNQGDNRRE 579
Db 506 SVNCELYFMVGVNSRTNTPVETWKGSKGQSYTIIIEENTTTSTWAFQRTTFHEASRKY 565
QY 580 INDMVKIYSITATNAVDGVASSCRACALGSEQSSCVPCCPPGHYIEKETNQCKECPDPT 639
Db 566 TNDVAKIYSINVTNMVNGVASYCRPCALEASDVGSSTSCPAGYIIDRDSGTCHSCPNT 625
QY 640 YLSIHQVYGEACIPCGSGSKNNQDHSVCYSDCFYHEKENQILHYDFSNLSSVGLMNG 699
Db 626 ILKAHQPYGVQACVPCPGTKNNKIHSICYNDCTFSRNTPTRTFNYSFALANTVTLAGG 685
QY 700 PSFTSKGTKYFFHFNISLCGHEGKMKALCTNNITDFTVKEIVAGSDDYTNLVGAFVCQST 759
Db 686 PSFTSKGLKYFHHTLSLGCNGQGRKMSVCTDNVTDLRIPE --GESGFSKSITAYVQAV 742
QY 760 IIPSESKGFRAALSSQSIIADTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSS 819
Db 743 IIPPEVTGYKAGVSSQPSVLADRLIGVTTDMTLDGITS PAELFHLESLGIPDVIFFYRSN 802
QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESAECPLCTEH 879
Db 803 DVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCDGTCDCGNFHLWESAACPLCSVA 862
QY 880 DFHEIEGACKRGFOETLYVWNEPKWCIKGISLPEKKLATCETVDFWLKVAGVGAFTAVL 939
Db 863 DYHAIVSSCVAGIQXTTYVXREP KLCSGGISLPEQRTVICKTIDFWLVKVISAGTCTAIL 922
QY 940 LVALTCYFWKKNQKLEYKSKLVMTTNSKECELPAAADSCAIMEGEDNEEEVYVSNKQSL 999
Db 923 LTVLTCYFWKKNQKLEYKSKLVMNATLKDCDLPAADSCAIMEGEDVEDDLIFTSKKSLF 982
QY 1000 GKLKSLATKEKEDHFESVOLKTS 1022
Db 983 GKIKSFTSKRTPDGFDSVPLKTS 1005
```

RESULT 11  
US-10-140-474-38  
; Sequence 38, Application US/10140474  
; Publication No. US20030032156A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc

```
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3330RIC162  
; CURRENT APPLICATION NUMBER: US/10/140,474  
; CURRENT FILING DATE: 2002-05-06  
; Prior Application removed - See Palm or File Wrapper  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 38  
; LENGTH: 1013  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 877, 882  
; OTHER INFORMATION: unknown amino acid  
US-10-140-474-38
```

```
Query Match 53.2%; Score 3025; DB 9; Length 1013;  
Best Local Similarity 54.4%; Pred. No. 1.3e-228;  
Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;  
  
QY 46 WAG---DLPPSSSRPLPPCOEKDYHFEYTECDSSSRWRVAIPNSAVDCSGLPDPVRGKE 102  
Db 31 WAGTAFQVQTGTGPELHACKESEYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKGT 90  
QY 103 CTFSASCAGEYLEMKNQVCSKCGEGTYSLSGSGIKFDEWDELPAFNSNIATFMDTVVGP 162  
Db 91 CSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIREDEWDELPHGFASLSANMELDDSAES 150  
QY 163 RPDGCNNSSWIPRGNYIESNRDDCTVSLIYAVHLKKSQYVFFEYQYVDNNIFFEFFFIOND 222  
Db 151 -TGNC TSSKWVPRGDYIASNTDECTATLMYAVNLKQSGTVNFEYYPDSSIIFFEFVQND 209  
QY 223 QCQEMD TTTDKWVLTONGEWSHVMKSGTNILYWR TTTGILMGSKAVKPVLVKNITIE 282  
Db 210 QCQP -NADDSRWMTTEKG -WEFHSVELNRGNVLYWRTTAFSVWTKVPKPVLRNIAIT 267  
QY 283 GVAYTSECFPCPGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCKDDSQFS --GSSECTE 340  
Db 268 GVAYTSECFPCPGTYADKQGSSEFCKLCPANSYSNKGGETSCHQC -DPDKYSEKSSSCNV 326  
QY 341 RPPCTTKDYFQIHTPCDEEGKTQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPGE 400  
Db 327 RPACTDKDYFYTHTACDANGETQLMYKWKAPKICSEDELEGAVKLPASGVKTHCPCPNPGE 386  
QY 401 YNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFYKWWNVLPGNMKTSCFNVGNSKC 460  
Db 387 FKTNNSTCQPCPYGSYNGS -DCTRCPAGTEPAVGFEYKWWNTLPTNMETTVLSGINFEY 445  
QY 461 DGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGKPPTS -MTGATGSELGRITVFETLC 519  
Db 446 KGMTGWEVAGDHIYTAAGASDNDFMILTLLVVPGRPPQSVMDATENKEVARITVFETLC 505  
QY 520 SADCVLVFMVDINRKSTNVVSWGGTKEKQAYTHIIFKNATFTFTWAFQRTNQGDNRRE 579  
Db 506 SVNCELYFMVGVNSRTNTPVETWKGSKGQSYTIIIEENTTTSTWAFQRTTFHEASRKY 565  
QY 580 INDMVKIYSITATNAVDGVASSCRACALGSEQSSCVPCCPPGHYIEKETNQCKECPDPT 639  
Db 566 TNDVAKIYSINVTNMVNGVASYCRPCALEASDVGSSTSCPAGYIIDRDSGTCHSCPNT 625
```

QY 640 YLSIHQVYGKEACIPCGPGSKNNQDHSVCSYDCFFYHEKENQILHYDFSNLSSVSGSLMNG 699  
Db 626 ILKAHQPYGVQACVPCGPGTKNNKTHSLCYNDCTFSRNTPTFTFNFSALANTVTLAGG 685  
QY 700 PSETSGTKYFHHFNISLCGHEGKMKALCTNNITDFTVKEIVAGSDDYTNLVGAFCQST 759  
Db 686 PSFTSKGLKYFHHFTLSLCGNQGRKMSVCTDNVTDLRIPE---GESGFSKSITAYVCOAV 742  
QY 760 IIPSESKGFRAALSSQSIIILADTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSS 819  
Db 743 IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSPAELFHLESGLIPDVIFFYRSN 802  
QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGTCDCGCTFYFLWESAEACPLCTEH 879  
Db 803 DVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCSDGTCDCGNFHLWESAAACPLCSVA 862  
QY 880 DFHEIEGACKRGFQETLYVWNEPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAFTAVL 939  
Db 863 DYHAIVSSCVAGIQXTTYVXREPKLCSGGISLPEQRVTICKTIDFWLKVGISAGTCTAIL 922  
QY 940 LVALTCYFWKKNQKLEYKYSKLVMTNTNSKECELPAADSCAIMEGEDNEEVVYSNKQSLL 999  
Db 923 LTVLTCYFWKKNQKLEYKYSKLVMTNATLKDCDLPAADSCAIMEGEDVEDDLIFTSKKSFL 982  
QY 1000 GKLKSLATKEKEDHFESVQLKTS 1022  
Db 983 GKIKSFTSKRTPDGFDSVPLKTS 1005

RESULT 12

US-10-142-431-38  
; Sequence 38, Application US/10142431  
; Publication No. US20030036179A1

GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C251

; CURRENT APPLICATION NUMBER: US/10/142,431

; CURRENT FILING DATE: 2002-05-10

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 38

; LENGTH: 1013

; TYPE: PRT

; ORGANISM: Homo Sapien

; FEATURE:

; NAME/KEY: unsure

; LOCATION: 877, 882

; OTHER INFORMATION: unknown amino acid

US-10-142-431-38

Query Match

Best Local Similarity 53.2%; Score 3025; DB 9; Length 1013;

Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;

QY 46 WAG-- -DLPSSSSRLPPCQEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKE 102  
Db 31 WAGTAFQVTOGTGPPELHACKESYHYEYTACDSTGSRWRVAVPHTPGLCTSLSDPVKGTE 90  
QY 103 CTESCASGEYLEMKNQVCSKCEGTYSLSGIKFDEWDELPAGFSNIAFMDTVVGPSDS 162  
Db 91 CSFSCNAGEFLDMKDQSCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAES 150  
QY 163 RPDGCNNSSWIPRGNYESNRDDCTVSLIYAVHLKKSGYVFFEQYQYVDNNIFFEQQND 222  
Db 151 -TGNCSTSSKWVPRGDYIASNTDECTATLMYAVNLKQSGTVNFEYYPDSSIIFFEQND 209  
QY 223 QCOEMDTTDDKWKLTNDGEGSHSVMLKSGTNILYWRITGILMGSKAVKPVLVKNITIE 282  
Db 210 QCQP-NADDSRWMTTEKG-WEFHSVELNRGNVLYWRTAFSVWTKVPKPVLRNIAT 267  
QY 283 GVAYTSECFPCPKPGTFESNKGFCNQCVCPRNTYSEKAKECIRCKDDSQFS--GSSECTE 340  
Db 268 GVAYTSECFPCPKPGTYADKQSSFCCLKCPANSYSNKGGETSCHQC-DPDKYSEKGSSSCNV 326  
QY 341 RPCTTKDYFQIHTPCDEEGKTQIMYKWIIEPKICREDLTDAILPPSGEKKDCPPCNP 400  
Db 327 RPACTDKDYFYHTACDANGETQLMYKWKAPKICSEDEGAVKLPASGVKTHCPCPCNP 386  
QY 401 YNNGSSSCHPCPPGTFSDGTECRPCPAGTEPALGFEYKWNVLPGNMKTSFCFNVGN 460  
Db 387 FKTNNSTCQPCPYGSYNGS-DCTRCPAGTEPAVGFEYKWNLTPTNMETTIVLSGINFEY 445  
QY 461 DGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTS-MTGATGSELGRITFVFETLC 519  
Db 446 KGMTGWEVAGDHIYTAAGASDNDFMILTLLVPGFRPPQSVMAADTENKEVARITFVETLC 505  
QY 520 SADCVLVFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTTFAFORTNQODNRRF 579  
Db 506 SVNCELYFMVGVSRTNTPVETWKGSKGQSYTYIIENNTTSFTWAFQRTTFHEASRKY 565  
QY 580 INDMVKIYSITATNAVDGVASSCRACALGSEQSGSSVCPCPPGPHYIEKETNOCEKPPDT 639  
Db 566 TNDVAKIYSINVTNVMNGVASYCRPCALEASDVGSSCTSCPAGYIIDRDSGTCHSCPNT 625  
QY 640 YLSIHQVYGKEACIPCGPGSKNNQDHSVCSYDCFFYHEKENQILHYDFSNLSSVSGSLMNG 699  
Db 626 ILKAHQPYGVQACVPCGPGTKNNKIHSLCYNDCTFSRNTPTFTFNFSALANTVTLAGG 685  
QY 700 PSFTSKGTKYFHHFNISLCGHEGKMKALCTNNITDFTVKEIVAGSDDYTNLVGAFCQST 759  
Db 686 PSFTSKGLKYFHHFTLSLCGNQGRKMSVCTDNVTDLRIPE---GESGFSKSITAYVCOAV 742  
QY 760 IIPSESKGFRAALSSQSIIILADTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSS 819  
Db 743 IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSPAELFHLESGLIPDVIFFYRSN 802  
QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGTCDCGCTFYFLWESAEACPLCTEH 879  
Db 803 DVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCSDGTCDCGNFHLWESAAACPLCSVA 862  
QY 880 DFHEIEGACKRGFQETLYVWNEPKWCIKGISLPEKKLATCETVDFWLKVGAGVGAFTAVL 939  
Db 863 DYHAIVSSCVAGIQXTTYVXREPKLCSGGISLPEQRVTICKTIDFWLKVGISAGTCTAIL 922  
QY 940 LVALTCYFWKKNQKLEYKYSKLVMTNTNSKECELPAADSCAIMEGEDNEEVVYSNKQSLL 999  
Db 923 LTVLTCYFWKKNQKLEYKYSKLVMTNATLKDCDLPAADSCAIMEGEDVEDDLIFTSKKSFL 982  
QY 1000 GKLKSLATKEKEDHFESVQLKTS 1022  
Db 983 GKIKSFTSKRTPDGFDSVPLKTS 1005

RESULT 13

US-10-143-114-38

; Sequence 38, Application US/10143114

; Publication No. US20030036180A1









Db	446	KGMTGWEVAGDHIYTAAGASDNDNFILTLVVPGRPPQSVMDATENKEVARITFVFETLC	505
QY	520	SADCVLYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTWAFQRTNQODNRRF	579
Db	506	SVNCELYFMVGVNSRTNPVETWKGSKGQSYTYIIEENTTTTSETWAFQRTTFHEASRKY	565
QY	580	INDMVKIYSITATNAVDGVASSCRACALGSEQSGSSVCPGPPGHYIEKETNOCKECPDPT	639
Db	566	TNDVAKIYSINVTNVMNGVASCRPCALEASDVSSCTSCPAGYYIDRDSGTCHSCPNT	625
QY	640	YLSIHQVYGKEACIPCGPGSKNNQDHSVYSDCFFYHEKENQILHYDEFNLSSVGSMLNG	699
Db	626	ILKAHPYGVQACVPCGPGTKNNKIHSLYNDCTFSRNTPTRTFNYNFSALANTVTLAGG	685
QY	700	PSFTSGTKYFHFNFISLCGHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFVCOST	759
Db	686	PSETSKGLKYFHHFTLSLCGNQGRKMSVCTDNVTDLRIPE---GESGFSKSITAYVQAV	742
QY	760	IIPSESKGFRAALSSQSIILADTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSS	819
Db	743	IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSPAELFHLESGLGIPDVIFVRSN	802
QY	820	TATTSINGRSTAVKMRCPNPKSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEH	879
Db	803	DVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPGTCSDDGTCDGCNHFHFWESAAACPLCSVA	862
QY	880	DFHEIEGACKRGFQETLYVWNEPKWCIKGISLPEKKLATCETVDFWLKVAGVGAFVAVL	939
Db	863	DYHAIVSSCVAGIQXTTYVXREPKLCSGGISLPEQRVTICKTIDFWLKVGISAGTCTAIL	922
QY	940	LVALTCYFWKKNQKLEYKYSKLVMTNSKECELPAAADSCAIMEGEDNEEEVYSNKQSLL	999
Db	923	LTVLTCYFWKKNQKLEYKYSKLVNATLKDCLPAAADSCAIMEGEDVEDDLIFTSKRSLE	982
QY	1000	GKLSLATKEKEDHFESVQLKTS	1022
Db	983	GKIKSFTSKRTPDGFDSVPLKTS	1005

Search completed: May 12, 2003, 13:37:33  
Job time : 191.369 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 11:39:43 ; Search time 29.9327 Seconds  
(without alignments)  
3298.405 Million cell updates/sec

Title: US-10-073-333A-4  
Perfect score: 5681  
Sequence: 1 MLFRARGPVRGRGWGRPAEA.....KEKEDHFESVQLKTSRSPNI 1027

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	242.5	4.3	1548	2 S34583	serine proteinase
2	210	3.7	713	2 A35502	major surface-labe
3	200	3.5	3635	2 T10053	laminin alpha 5 ch
4	199.5	3.5	2823	2 T23064	hypothetical prote
5	199.5	3.5	2823	2 F87908	protein T22A3.8 [i
6	199.5	3.5	3102	2 T43291	laminin alpha chai
7	198.5	3.5	677	2 C42125	trophozoite cystei
8	194	3.4	5376	2 T42215	zonadhesin - mouse
9	186	3.3	1766	2 A42125	trophozoite cystei
10	183	3.2	1372	2 T25933	hypothetical prote
11	182	3.2	667	2 A48579	trophozoite surfac
12	181	3.2	1790	1 MMFFB1	laminin beta-1 cha
13	178.5	3.1	3672	2 T23433	hypothetical prote
14	178.5	3.1	3704	2 T37316	probable laminin a
15	177.5	3.1	1607	1 MMMSB2	laminin gamma-1 ch
16	177.5	3.1	2907	2 A57278	fibrillin-2 precur
17	174.5	3.1	596	2 A45664	variant-specific s
18	173.5	3.1	2871	2 A55624	fibrillin-1 precur
19	173	3.0	1111	2 T26972	hypothetical prote
20	173	3.0	1786	1 MMMSB1	laminin beta-1 cha
21	171	3.0	1620	2 T27283	hypothetical prote
22	170.5	3.0	2437	2 S42612	transmembrane prot
23	170	3.0	1557	2 T28811	hypothetical prote
24	170	3.0	1786	1 MMHUB1	laminin beta-1 cha
25	166.5	2.9	557	2 A48434	variant-specific s
26	166.5	2.9	1274	2 T42017	cysteine rich prot
27	166	2.9	3002	2 A47221	fibrillin 1 precur
28	165	2.9	1639	1 MMFFB2	laminin gamma-1 ch
29	164.5	2.9	2918	2 A54105	fibrillin-2 precur

30	164	2.9	2491	1 A28372	insulin-like growt
31	164	2.9	2871	2 A55567	fibrillin I - bovi
32	164	2.9	3084	1 MMMSA	laminin alpha-1 ch
33	163	2.9	1297	2 T30274	proteoliasin - se
34	161	2.8	3712	2 S18253	laminin alpha-1 ch
35	160.5	2.8	1609	1 MMHUB2	laminin gamma-1 ch
36	160	2.8	1220	2 A56136	jagged protein pre
37	159.5	2.8	2352	2 T30201	Notch homolog prot
38	157	2.8	2531	2 A46019	Notch-1 protein -
39	156.5	2.8	3106	1 S53868	laminin alpha-2 ch
40	155	2.7	2499	1 A30788	mannose 6-phospat
41	154.5	2.7	13288	2 T03099	mucin, submaxillar
42	154	2.7	1323	2 PN0568	connectin 3B - chi
43	153.5	2.7	1680	2 A43434	furin (EC 3.4.21.7
44	153	2.7	1959	1 AGRT	agrin - rat
45	153	2.7	2318	2 S45306	notch 3 protein -

ALIGNMENTS

RESULT 1  
S34583

serine proteinase (EC 3.4.21.-) PC6B - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 02-Dec-1993 #sequence\_revision 10-Nov-1995 #text\_change 05-Nov-1999  
C;Accession: S34583  
R;Nakagawa, T.; Murakami, K.; Nakayama, K.  
FEBS Lett. 327, 165-171, 1993  
A;Title: Identification of an isoform with an extremely large Cys-rich region of PC6,  
A;Reference number: S34583; MUID:93327934; PMID:8335106  
A;Accession: S34583

A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-1548 <NAK>

A;Cross-references: GB:D17583; NID:g407344; PIDN:BAA04507.1; PID:d1005033; PID:g44037  
C;Keywords: hydrolase; serine proteinase

Query Match 4.3%; Score 242.5; DB 2; Length 1548;  
Best Local Similarity 18.4%; Pred. No. 5.9e-08;  
Matches 248; Conservative 116; Mismatches 432; Indels 549; Gaps 72;

QY	61	CQEKDYHFYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKEC-----TFSCA	108
Db	347	CPPGHYHADKKRC-----RKCAPN-CESCFG----SHGQCLSKYGYFLNEETSSCV	394
QY	109	S-----GEYLEMKNQVCCKGEGTYSLGSIGKFDWEDELPAFES-----NIATFMDTVVG	158
Db	395	TQCDGSYEDIKKNVCGKCSENCAC---IGFHNCTECKGLSLQGSRCSVTCEDGQFFN	451
QY	159	PSDSRP-----DGCNNSWIPRGNYESNR--DDCTVSLIYAVHLKKSGYVF	203
Db	452	GHDCQPCHRFCATCSGAGADGCINCT---EGYVMEEGRCVQSCSVS--YYLDHSSEGGYK-	506
QY	204	FEQYVDNNIFF-----EFFIQNDQCQEMDTTT	231
Db	507	-SCKRCDNSCLTCNGPGFKNCSSCPGYLLDLGTCQMGAI CKDGEYIDDQGH CQTCEASC	565
QY	232	DKWVKLTDNGEGWSHVMKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECF	291
Db	566	AKC-----WG-----PTQEDCISCPVTRVLDDGRC-----VMNCPSWKFFFKQCH	606
QY	292	PCKPGTFSNKPGSFNCQVCPRNTYSEKGAKECIRCDD-----SQFSGSSECTERPPCTTK	347
Db	607	PC-----HYTCQGC-----QSGGPSNCTSCRADKHGQERFLYHGECLEN--CPVG	649
QY	348	DY-FQIHT--PCDEEGKTQIMYKWIIEPKICREDLTDAIRLPPS-----GE	389
Db	650	HYPAKGHTCLPCPD--NCELCY---NPHICSRMSGYVIIPPNTHTCQKLECRQGEFQDSE	704
QY	390	KKDCPPCPNPGFY---NNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFYKWNVLPG	446
Db	705	YEEMPCEEGLGCTEDDPGACTSCATGYMFERHCYKACPEKT-----FGVKWEACRG	759

QY 447 NMKTSFNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGS 506  
Db 760 ---TNGSCDQHEC---YWCEGFFLGGSCVQDCG-----PGFH-----GDO 796  
QY 507 ELGRITFVFETLCSADCVLVFMVDINRKST--NVVESWGKTKEKQAYTHIIFKNATFTFT 564  
Db 797 ELGEC-----KPCRACETCTGSGYNQCSSCQEGQLWHGTCLWSTWPOVEGKD----- 845  
QY 565 WAFQRTNQG-----QDNRRFINDMVKIYSITATNAVDGVASSCRACALGSEQ 611  
Db 846 W-----NEAVPTEKPSLVRSLQDRRKW---KVQI-KRDATSONQPCCHSSCKTC----- 890  
QY 612 SGSSCVPCPPGHI-----EKTNOCKEC-----PPD 638  
Db 891 NGSLCASCPTGMYLWLQACVPSPQGTWPSVTSGSCEKCEDCVSCSGADLCQOCLSQPD 950  
QY 639 TYLSIHQ-----VYGKE-ACIPCGPGSKNNQ-----DHSVCY--- 669  
Db 951 NTLLHHEGRCYHSCPEGFYAKDGVCEHCSSPCKTCEGNATSCNSCEGDFVLDHGVCKWTC 1010  
QY 670 -----SDCFFVHEK-----ENQILHYDFNSLSSVGL----- 696  
Db 1011 PEKHVAVEGVCKHCPCQDCI--HEKTCCEMPDFFLYNDMCHRSCPKSFYPDMRQCVP 1068  
QY 697 -----MNGPSFTSGTKYFHFFNISLCGHEGKKM--ALCTNNITDFTVKE----- 739  
Db 1069 CHKNCLCNGPKED-----DCKVCADTSKALHNLGLCLDECPEGTYKEEENDECRD 1118  
QY 740 ----IVAGSDDYTNLV-----GAFVCQSTIIPSESKGER 769  
Db 1119 CPESCLICSSAWTCLACREGFTVVHVDVCTAPKEAAVEYWDEGSHRCQ----PCHKKCSR 1174  
QY 770 AALSSQSILADTFIGTVVETTLKNIN- IKEDMFPVPTSQIPDVHFFYKSSATTSTC--- 825  
Db 1175 CSGPSE-----DQCYTCPRETFLNNTCVKE-----CPEGYHTDKDSQQCVLCHSS 1220  
QY 826 ---INGRSTAVKMRNPTKSGAGVISVPSKC-----PAGTCDGC----- 861  
Db 1221 CRTCEGPHSQCLSCR-----GWFLGKECLLQCRDGYGYGESTSGRCEKCDKSKSCRG 1275  
QY 862 -----TFYFLWESAEAC-PLCTEHDF-----HEIEGAC--KRGFQETLY 897  
Db 1276 PRPTDCQSCDTFFLLRSKGQCHRRACEHYAYADQHAQTCERCHPTCDKCSGKEAWSCLSC 1335  
QY 898 VVNEPKWCIKGISLPE-----KLLATCET----- 921  
Db 1336 VWS--YHLKGICICEIVGEYREGKGFENCKKCHESCKECPGSKNCTGCSAGLLLD 1393  
QY 922 -----VDFW-----LKVAGVGAFTAV 938  
Db 1394 MDDNRCLHCCNASHRRSQDCCQSCSTDECILPAREAEFYEHKTAALLVTSGA---MLL 1450  
QY 939 LLVALTCTYFKKNQK---LEYKYSKLVMTNS-KECELPAADSCAIME-----GEDNEE 988  
Db 1451 LLLGAAAVVWRKSRSPVAKGRYEKLAEPTVSYSSRYLDEQVIEYRDRDYDEDED 1510  
QY 989 EVVYSNKQ-SLLGKLKSLATKEKED 1012  
Db 1511 DIVYMGQDGTVYRKFKYGLLDETD 1535  
RESULT 2  
A35502  
major surface-labeled trophozoite antigen precursor - Giardia lamblia  
C:Species: Giardia lamblia  
C:Date: 09-Nov-1990 #sequence\_revision 09-Nov-1990 #text\_change 09-Sep-1997  
C:Accession: A35502  
R:Gilllin, F.D.; Hagblom, P.; Harwood, J.; Aley, S.B.; Reiner, D.S.; McCaffery, M.; So, M.  
Proc. Natl. Acad. Sci. U.S.A. 87, 4463-4467, 1990  
A:Title: Isolation and expression of the gene for a major surface protein of Giardia lam  
A:Reference number: A35502; MUID:90280395; PMID:2352929  
A:Accession: A35502

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-713 <GIL>  
A:Cross-references: GB:M33641; NID:gl59131; PID:gl59132  
C:Keywords: surface antigen; transmembrane protein  
Query Match 3.7%; Score 210; DB 2; Length 713;  
Best Local Similarity 19.7%; Pred. No. 3.4e-06;  
Matches 186; Conservative 96; Mismatches 300; Indels 360; Gaps 53;  
QY 90 DCSGLP-DPVRGKECTFSCASGEYLEMKNQ---VCSKGG-----EGTSLGSGIK-- 135  
Db 43 ECGANYAPVNG-QCVDVNAEGPSKTLCPQHSAGKCTQCGNSFMYKDKGCYSSGEGLPGH 101  
QY 136 ----FDEWDEL-----PAGEFSN-----IATEMDTVVGPSPD---S 162  
Db 102 SLCLSSDGDGVCTEAPGYFAPVGAANTEQSVIACGDTTGTIAAGNTYKGIADCAECS 161  
QY 163 RPDGNNSSWIPRGNIESNRDDCTVSLIYAVHLKKSGYVFFFEYQYVDNNIFFEFTQND 222  
Db 162 APDA---TAGAEAGKVATCTK---CGVS-----KYLKDNVCV-----D 193  
QY 223 QCQEMDTTDDKWKLTNDGEWGSHSVMLKSGTNILYWRITGILMGSKAVKPVLVKNITIE 282  
Db 194 KAQNSGSGTNKEVAV-DDSENGNKCVCSDNLN-----G 226  
QY 283 GVAYTSECFCKPGTFSNKPSPNCQVCPRTYSEKGA-ECIRCKDDSQSGSSE---C 338  
Db 227 GVA-----NCDTC---SYDEQSKIKCTKCTDNNYLTSTSEGTSC 263  
QY 339 TERPCTTKDYFQIHTPCDEEGTKIIMYKWIETKICREDLTDAILPPSGEKKDCPPCNP 398  
Db 264 VQKQDC--KDGf--FPKDDSS-----AGNK--CLPCND 290  
QY 399 GFYNNSSSCHPCP--PGTFSDGTECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVG 456  
Db 291 S--TDGIANCATCALVSGRGAALVTCASCTDGYKPSAD-----KTTCEAVS 335  
QY 457 NSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITFVFE 516  
Db 336 NCKTPCKKACSNEGKENEVCTDCDGSYTL-----TPTSQ----- 369  
QY 517 TLCSDCVLYFMVDINRKSTNVVSWGGTKEKQAYTHIIFKNATFTFTWAFQRT--NQOQ 574  
Db 370 --CIDSQA-----KIGNYYGATEGAKK-----LCKECTA-----ANCKTCDQDQ 407  
QY 575 DNRRTINDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHIYIEK-----ET 629  
Db 408 -----CQACNDGFYKNGDACSPC---HESCKTCSAGTA 437  
QY 630 NOCKECPDPTYLSIHQVYKAEACI-PCGPGSKNNQDHSVCYSDCFYHEKENQILHYDFS 688  
Db 438 SDCTECPGKALR---YGDGTPKTCGEGCTTGTGAGAC----- 473  
QY 689 NLSSVGLMNGPSTSGTKYFHFFNISLCGHEGKMAICTNNITDFTVKEIVAG--SDD 746  
Db 474 --KTCGLTIDGASYCSECATTTTEYPONGVC---APKASRATPTCNDSPIONGVCGTCA 528  
QY 747 YTNLVGAFVCQSTIIPSESKGFRAALSSQSILADTFIGTVVETTLKNINIKEDMFPVPT 806  
Db 529 YFKMNGG--CYETV---KYPGKTVCTISAPN-----GGTCOKAADGYKLDGSLTVCS 575  
QY 807 SQIPDVHFFYKSSSTATTSING--RSTAVKMRN---PTKSGAGVISVPSKCPAGTCDGC 861  
Db 576 EGCKEC-----ASSTDCTTCLDGYVKSASACTKCDASCETCNGA-----ATTCKAC 621  
QY 862 -TFYFLWESAE-ACPLCTEHDFHEIEGACKRGFQETLYVWNEPKWIKGISLPEKK-LAT 918  
Db 622 ATGYKYKTASGEAGCTSC-ESDSNGVTGI-----KGCINCAAPPNNKGSVL 665  
QY 919 C-----ETVDFWLKVGAGVGAFTAVL-----LVALTCYFW 948  
Db 666 CYLIKDSGSGTNKSGSLSTGAIGISVAVIVVVGGLIGFLCWWF 707





QY 301 --KPGSFN-----CQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCT-----TKD 348  
Db 1020 QCRPSYFNFTDAGCQFCHCNIY---GSIEDGKC---DQTTGKCECENVEGTMCEKCADG 1073  
QY 349 YFQIHT-----PCDEEGKTQIMYKWIPEK-ICREDLTDAIRLPPSGEKKDCPPCPNGF 400  
Db 1074 YFNITSGDGEDCGCDPTGSEDVSCNLVTGQCVCKPGVT-----GLK--CDSCLPNF 1123  
QY 401 Y---NNGSSSCHPCP-PGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVG 456  
Db 1124 YGLTSEGCTECPAPGVCDPIDGSCVCPNPT-----VEMCENCNTT-- 1167  
QY 457 NSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITFVFE 516  
Db 1168 -----NAW-----DYHPLN-----GCK----- 1179  
QY 517 TLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAY-----THIIFKNATFTTWAFO 568  
Db 1180 -LC--DC-----SDIGSDGGMCMNTFTGQCKCKAAYVGLKCDLCTHGFFNFPT----- 1223  
QY 569 RTNOGQDNRFRINDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKE 628  
Db 1224 -----CEPC--GCNAAGTDPLQCKDGQCL--- 1245  
QY 629 TNQCKECPDPTYLSIHQVYGKEACIPCGPG-----SKNNQDHSVYSDCF----- 673  
Db 1246 CNEIGECP---CKKNVHGK-CDQCGEGTFLDSSNLKGCTECF--CFNRTSNCEQSD 1297  
QY 674 -----FYHEKEN-QILHYDFSNLSS-----VGLSLMNGPS 701  
Db 1298 LVWQQMYAEDRRVAFQEPWEFYTKKHNNILLREKPSHFNSYPTDATPLYWPLPSTMLGDR 1357  
QY 702 FTS-----KGTKYF-HFFNISLCGHEGKKMALCTNNITDFTVK 738  
Db 1358 TASYNGFLRFKIWNEDNRRLHGIIRDQOQYFRHFPQVVIIFGNRIELEHIPMEINDDGIY 1417  
QY 739 EI-----VAGSDDYT-----NLVGAFCVQSTIIPSESKGFRAALSSQSIIIL 779  
Db 1418 KIRLHESERVRHSPELTLTRKQMMVALQDTQGIYIRGTYYTPARGDAINIOEVSLDVAV 1477  
QY 780 ADTFIGTVVETTLKNINIKEDM-----FPVPTSQIPDVHFFYKSS----- 819  
Db 1478 PESKIVAGLSTT-KAIGVEKCLGCPQGYTGLSCQNPVEGVYRKKHREYLNQADDIALIGW 1536  
QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGT---CDGCTFYFLWESAE--- 871  
Db 1537 SEPCSC-HGHS---QTCNPD-----TSVCTDCEHNTFGDFCEHCLPGYIGDAREGGAN 1585  
QY 872 -----ACPL-----CTEHDFHEIEGACKRGFQETLYVWNEPKWCIKG-ISLPEKK 915  
Db 1586 ACTKCACPLVENSFSDSCVAVDHGRGYVGNCKPG-----YTQYQETCVAGYIGDPQHI 1640  
QY 916 LATCETVDFWLKVAGVGAFATVL-----LVALTCYFWKKNKQLEKYKSLVMTTNS 967  
Db 1641 GGTCSPCDCH-PDGLSLHGACNPLSGQCECKPGVTGRTCSMCQE----RHAFINRVCTSCD 1695  
QY 968 KECELPAADSCAIMEGEDNEEVVYSNKQSLGLKLSLATK 1008  
Db 1696 QGCYLP-----LMETMDTMEE--HLGRONFSG-LKPIPWK 1727  
RESULT 5  
F87908  
protein T22A3.8 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-Aug-2001  
C:Accession: F87908; E87908  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_elegans  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A;Accession: F87908  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2823 <STO>  
A;Cross-references: GB:chr\_I; PIDN:CAAL5432.1; PID:g3924779; GSPDB:GN00019; CESP:T22A  
A;Accession: E87908  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2823 <ST2>  
A;Cross-references: GB:chr\_I; PIDN:CAB03385.1; PID:g3924881; GSPDB:GN00019; CESP:T22A  
C;Genetics:  
A;Gene: T22A3.8  
A;Map position: 1  
C;Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like

Query Match 3.5%; Score 199.5; DB 2; Length 2823;  
Best Local Similarity 19.1%; Pred. No. 9.3e-05;  
Matches 225; Conservative 110; Mismatches 363; Indels 483; Gaps 65;  
QY 39 LAGQAAWAGDLPSSSRPLPPC-----QEKDY-----HFEYTECDSGS 78  
Db 819 LGACEQC---ECPSLDLNPNPECISTELAVLGVSASVEDNVVCINCLPLYEGNKCEY--- 872  
QY 79 RWRVAIPNSAVDCSG--LPDPVRGK--ECTF-----SCASGEYLEMKNQVSKC- 123  
Db 873 -----CSDGFFEDPLTGKIECTCNGNIDPMGIGNCDS-----ETGKCLKCI 914  
QY 124 GEGTSLGSGGIKFDWEDELPAFNSNIATFMDTVVGPDSRDPDGCNNSSWIPRGNYIESNR 183  
Db 915 GHTTGDSCESCKEHHW-----GNAQLHT-----CKPCGCHTQGA VNPQCSEENGE 959  
QY 184 DDCTVSLIYAVHLKKSGYVFFEYQYVDNNIFFEFFFQNDQCEMDTTTDKWKVLTNDGEW 243  
Db 960 CEKENYIGA-----QCDRCKENHGDVENGCPACDCNDT 993  
QY 244 GSHSVMLKSGTNILYWRTTGILMGSKAVKPVLVKNITIEGVAYTSECPCPKPGTFSN--- 300  
Db 994 GS-----IGSDCDQ-----VSGQC-NCKQGVFGKQCD 1019  
QY 301 --KPGSFN---CQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCT-----TKD 348  
Db 1020 QCRPSYFNFTDAGCQFCHCNIY---GSIEDGKC---DQTTGKCECENVEGTMCEKCADG 1073  
QY 349 YFQIHT-----PCDEEGKTQIMYKWIPEK-ICREDLTDAIRLPPSGEKKDCPPCPNGF 400  
Db 1074 YFNITSGDGEDCGCDPTGSEDVSCNLVTGQCVCKPGVT-----GLK--CDSCLPNF 1123  
QY 401 Y---NNGSSSCHPCP-PGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVG 456  
Db 1124 YGLTSEGCTECPAPGVCDPIDGSCVCPNPT-----VEMCENCNTT-- 1167  
QY 457 NSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITFVFE 516  
Db 1168 -----NAW-----DYHPLN-----GCK----- 1179  
QY 517 TLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAY-----THIIFKNATFTTWAFO 568  
Db 1180 -LC--DC-----SDIGSDGGMCMNTFTGQCKCKAAYVGLKCDLCTHGFFNFPT----- 1223  
QY 569 RTNOGQDNRFRINDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEKE 628  
Db 1224 -----CEPC--GCNAAGTDPLQCKDGQCL--- 1245  
QY 629 TNQCKECPDPTYLSIHQVYGKEACIPCGPG-----SKNNQDHSVYSDCF----- 673  
Db 1246 CNEIGECP---CKKNVHGK-CDQCGEGTFLDSSNLKGCTECF--CFNRTSNCEQSD 1297  
QY 674 -----FYHEKEN-QILHYDFSNLSS-----VGLSLMNGPS 701  
Db 1298 LVWQQMYAEDRRVAFQEPWEFYTKKHNNILLREKPSHFNSYPTDATPLYWPLPSTMLGDR 1357  
QY 702 FTS-----KGTKYF-HFFNISLCGHEGKKMALCTNNITDFTVK 738  
Db 1358 TASYNGFLRFKIWNEDNRRLHGIIRDQOQYFRHFPQVVIIFGNRIELEHIPMEINDDGIY 1417  
QY 739 EI-----VAGSDDYT-----NLVGAFCVQSTIIPSESKGFRAALSSQSIIIL 779  
Db 1418 KIRLHESERVRHSPELTLTRKQMMVALQDTQGIYIRGTYYTPARGDAINIOEVSLDVAV 1477  
QY 780 ADTFIGTVVETTLKNINIKEDM-----FPVPTSQIPDVHFFYKSS----- 819  
Db 1478 PESKIVAGLSTT-KAIGVEKCLGCPQGYTGLSCQNPVEGVYRKKHREYLNQADDIALIGW 1536  
QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGT---CDGCTFYFLWESAE--- 871  
Db 1537 SEPCSC-HGHS---QTCNPD-----TSVCTDCEHNTFGDFCEHCLPGYIGDAREGGAN 1585  
QY 872 -----ACPL-----CTEHDFHEIEGACKRGFQETLYVWNEPKWCIKG-ISLPEKK 915  
Db 1586 ACTKCACPLVENSFSDSCVAVDHGRGYVGNCKPG-----YTQYQETCVAGYIGDPQHI 1640  
QY 916 LATCETVDFWLKVAGVGAFATVL-----LVALTCYFWKKNKQLEKYKSLVMTTNS 967  
Db 1641 GGTCSPCDCH-PDGLSLHGACNPLSGQCECKPGVTGRTCSMCQE----RHAFINRVCTSCD 1695  
QY 968 KECELPAADSCAIMEGEDNEEVVYSNKQSLGLKLSLATK 1008  
Db 1696 QGCYLP-----LMETMDTMEE--HLGRONFSG-LKPIPWK 1727

Db 1358 TASYNGFLRFKIWNEDNRRGLHGIRPDQYFRHFPOVLIIFGNRRIELEHIPMEINDDGIY 1417

QY 739 EI-----VAGSDDYT-----NLVGAFCVQSTIIPSESKGFRAALSSQSIIIL 779

Db 1418 KIRLHESEWRVRHSPELTLTRKOMMVALQDTQGIYIRGTYYTPARGDAINIQEVSLDVAV 1477

QY 780 ADTFIGVTETTLKNINIKEDM-----FPVPTSQIPDVHFFYKSS----- 819

Db 1478 PESKIVAGLSTT-KAIGVEKCLGCPQGYTGLSCQNPEVGYRKKHREYLNQADDIALIGW 1536

QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGT-----CDGCTFYFLWESAE---- 871

Db 1537 SEPCSC-HGHS-----QTCNPD-----TSVCTDCEHNTFGDFCEHCLPGYIGDAREGGAN 1585

QY 872 -----ACPL-----CTEHDFHEIEGACKRGFQETLYVWNEPKWCIGK-ISLPEKK 915

Db 1586 ACTKCACPLVENSFSDSCVAVDHGRGYVNCCKPG-----YTGQYCETCVAGYYGDPQHI 1640

QY 916 LATCETVDFWLKVGAGVGAFVAVL-----LVALTCYFWKKNQKLEYKYSKLVMTTNS 967

Db 1641 GGTCSPCDCH-PDGLSLHGACNPLSGQCECKPGVTGRTCSMCQE----RHAFINRVCTSCD 1695

QY 968 KECELPAADSCAIMEGEDNEEVVYSNKQSLGLKLSLATK 1008

Db 1696 QGCYLP-----LMETMDTMEE--HLGRQNFSG-LKPIPWK 1727

RESULT 6

T43291

laminin alpha chain - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 04-Mar-2000

C:Accession: T43291

R:Zhu, X.; Kao, G.; Joh, K.; Hall, D.H.; Wadsworth, V.K.; Hutter, H.; Vogel, B.E.; Huang

submitted to the EMBL Data Library, June 1998

A:Description: Expression, function and evolution of laminin alpha chains.

A:Reference number: 222397

A:Accession: T43291

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-3102 <ZHU>

A:Cross-references: EMBL:AF074902; PIDN:AAC26793.1

C:Genetics:

A:Map position: 1

A>Note: lamal/2

C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like

Query Match 3.5%; Score 199.5; DB 2; Length 3102;

Best Local Similarity 19.1%; Pred. No. 0.0001;

Matches 225; Conservative 110; Mismatches 363; Indels 483; Gaps 65;

QY 39 LAGQAAWAGDLPSSSSRPLPPC-----QEKDY-----HFEYTECDSSGS 78

Db 819 LGACEQC--ECPSLDLNPPECISTELAVLGSVASNEDNVVINCPLGYEGNKCEY--- 872

QY 79 RWRVAIPNSAVDCSG--LPDPVRGK--ECTF-----SCASGEYLEMKNQVCSKC- 123

Db 873 -----CSDGFFEDPLTGKCIECTCNGNIDPMGIGNCDS-----ETGKCLKCI 914

QY 124 GEGTYSLGSGIKFDEWDELPAFGFSNIATFMDTVVGPDSRPDGCNNSWIPRGNVIESNR 183

Db 915 GHTTGDSCESCKEHHW-----GNAQLHT-----CKPGCCHTQGA VNPQCSENGE 959

QY 184 DDCVTSLIYAVHLKKSQYVFFEYQYVDNNIFFEFFFIONDQCQEMDTTDDKWVCLTDNGEW 243

Db 960 CEKENYIGA-----QCDRCKENHGDVENGCPACDCNDT 993

QY 244 GSHVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECFCKPGTFSEN--- 300

Db 994 GS-----IGSDCDQ-----VSGQC-NCKQGVFGKQCD 1019

QY 301 --KPGSFN-----CQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCT-----TKD 348

Db 1020 QCRPSYFNFTDAGCQFCHCNIY---GSIEDGKC---DQTTGKCECRENVEGTMCEKCADG 1073

QY 349 YFQIHT-----PCDEEGKTQIMYKWIPEK-ICREDLTDAIRLPPSGEKKDCPPNPGF 400

Db 1074 YFNITSGDGEDCGCDPTGSEDVSNLVTGQCVCKPGVT-----GLK--CDSCLPNF 1123

QY 401 Y---NNGSSSCHPCP-PGTFSDGTKECRPCPAGTEPALGFEYKWNVLPGNMKTSCFNVG 456

Db 1124 YGLTSEGCTECPAPGQVCDPIDGSCVCPNT-----VGEMCENCIT-- 1167

QY 457 NSKCDGMNGWEVAGDHIQSGAGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITFVFE 516

Db 1168 -----NAW-----DYHPLN-----GCK----- 1179

QY 517 TLCSADCVLYFMVDINRKSTNVVESWGGTKEKQAY-----THIIFKNATFTTTFWAFQ 568

Db 1180 -LC--DC-----SDIGSDGGMCTFTGQCKCKAAAYVGLKCDLCTHGFFNFPT----- 1223

QY 569 RTNOGQDNRRFINDMVKIYSITATNAVDGVASSCRACALGSEOSGSSCVPCPPGHYIEKE 628

Db 1224 -----CEPC--GCNAAAGTDPLQCKDQGQCL--- 1245

QY 629 TNOCKECPCPDYLSIHQVYGKEACIPCGPG-----SKNNQDHSVYCYSDCF----- 673

Db 1246 CNEICECP-----CKKNVHGTK-CDQCGEGTFSLDSSNLKGTGECF--CFNRTSNCEQSD 1297

QY 674 -----FYHEKEN-QILHYDFSNLSS-----VGLMNGPS 701

Db 1298 LVWQMYAEDRRRAVFQEPWEFYTKKHNNILLREKPSHFNSYPTDATPLYWPLPSTMLGDR 1357

QY 702 FTS-----KGTKYF-HFFNISLCGHEGKKMALCTNNITDFIVK 738

Db 1358 TASYNGFLRFKIWNEDNRRGLHGIRPDQYFRHFPOVLIIFGNRRIELEHIPMEINDDGIY 1417

QY 739 EI-----VAGSDDYT-----NLVGAFCVQSTIIPSESKGFRAALSSQSIIIL 779

Db 1418 KIRLHESEWRVRHSPELTLTRKOMMVALQDTQGIYIRGTYYTPARGDAINIQEVSLDVAV 1477

QY 780 ADTFIGVTETTLKNINIKEDM-----FPVPTSQIPDVHFFYKSS----- 819

Db 1478 PESKIVAGLSTT-KAIGVEKCLGCPQGYTGLSCQNPEVGYRKKHREYLNQADDIALIGW 1536

QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGT-----CDGCTFYFLWESAE---- 871

Db 1537 SEPCSC-HGHS-----QTCNPD-----TSVCTDCEHNTFGDFCEHCLPGYIGDAREGGAN 1585

QY 872 -----ACPL-----CTEHDFHEIEGACKRGFQETLYVWNEPKWCIGK-ISLPEKK 915

Db 1586 ACTKCACPLVENSFSDSCVAVDHGRGYVNCCKPG-----YTGQYCETCVAGYYGDPQHI 1640

QY 916 LATCETVDFWLKVGAGVGAFVAVL-----LVALTCYFWKKNQKLEYKYSKLVMTTNS 967

Db 1641 GGTCSPCDCH-PDGLSLHGACNPLSGQCECKPGVTGRTCSMCQE----RHAFINRVCTSCD 1695

QY 968 KECELPAADSCAIMEGEDNEEVVYSNKQSLGLKLSLATK 1008

Db 1696 QGCYLP-----LMETMDTMEE--HLGRQNFSG-LKPIPWK 1727

RESULT 7

C42125

trophozoite cysteine-rich surface antigen 72 - Giardia lamblia (fragment)

N;Alternate names: CRP72

C;Species: Giardia lamblia

C;Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 05-Dec-1998

C;Accession: C42125

R;Adam, R.D.; Yang, Y.M.; Nash, T.E.

Mol. Cell. Biol. 12, 1194-1201, 1992

A;Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170

A;Reference number: A42125; MUID:92186850; PMID:1545800

A;Accession: C42125

A;Molecule type: DNA

A;Residues: 1-677 <ADA>

A;Cross-references: GB:M83934; NID:g159123  
A;Experimental source: trophozoites  
A;Note: sequence extracted from NCBI backbone (NCBIN:88443, NCBIP:88444); this ORF is not  
C;Keywords: surface antigen

Query Match 3.5%; Score 198.5; DB 2; Length 677;  
Best Local Similarity 17.4%; Pred. No. 1.9e-05;  
Matches 181; Conservative 87; Mismatches 268; Indels 507; Gaps 47;

Qy 7 GPVGRGWGRPAEAPRRGRSPWPSPAWICWALAGCQAAWAGDLPSSSSRPLPPCQE-KD 65  
Db 35 GGVPVDGFCRPFQSPQA-----AAAGCTKAGGAALDKMTA-----TCEKCGD 76  
Qy 66 YHFEY-----TECDSSGSRNR--VAIPNSAVDC----- 91  
Db 77 GYFLFMGGCYKTTDGPGEICTKAEGGLCTECKTANGLFKNPAATPEKGSECILCSDING 136  
Qy 92 -SGLPDPVRGKECTFS-----CASGEYLEMKNQVCSKCGEGTYSGLSGIKFDEW 139  
Db 137 ADGYTGAVNCAOCTKSDSNKGAATCTACQAGYYKDF--QACSKC-DGT----- 181  
Qy 140 DELPAGFSNIATFMDTVVGPDSRDPG-----CNNSSWIPRGNYYIESNRDDCTVSLI 191  
Db 182 -----CLTCETSAAOCTSCPEGKYLKGDKSCVNNNGCTGNTYADPESGKC----- 226  
Qy 192 YAVHLKSGYVFPEYQYVDNNIFFEFFFIONDQCQEMDTTDDKWKVLTNDNGEWSHVMK 251  
Db 227 -----LPCNTIDQAC-----TQC-EVDSTTKK-PKCTNCG----- 254  
Qy 252 SGTNLYWRTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCPKGTFSNKPGSFNCQVCP 311  
Db 255 -----GQKMVK-----TAIDG---TTTCVDANGCATSNVDGSH----- 284  
Qy 312 RNTYSEKAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEGKTOIMYKWIEP 371  
Db 285 ---FLNDGSTRKILCSDDSLEANKGT--PGC----- 312  
Qy 372 KICREDLTDAIRLPPSGEKKDCPPCPGYNNGSSSCHPCPPCTFSDGTKECRPCPAGTE 431  
Db 313 KTCKK-----NGAKPTCSECLDGYNSGNG-----GTVTCEACGANC- 349  
Qy 432 PALGFEYKWNVLPGNMKTSFCNVGNSKC-----DGMNGWEVAGDHIQSG---- 476  
Db 350 -----ATCTQAGNDKCTKCKPGFFMKNGGTGECVACDNAQGGIDGC 391  
Qy 477 -----AGGSDNDYLILNLHIPPFGFKPPTSMGTATGSELGRITFV 514  
Db 392 AECTKESTGPLKCTKCKPNRKPAGTSDN-YTCTE---KTCENPTACGGTAGS----- 439  
Qy 515 FETLCSADCVLVFMVDINRKSTNVVESWGGTKEQAYTHIIFKNATFTTWFARTNQOQ 574  
Db 440 ----CDA-----IVIDDQGTCKHYCSYC-----GD 460  
Qy 575 DNRRFINDMKIYSITATNAVGVASSCRACALGSEQSGSSCVPCPPGHIYEKTNQCKE 634  
Db 461 SSQA-----PIDGL-----CASEAQKAGNTCA-----NGVCTQ 488  
Qy 635 CPPDTYLSIHQVYGKEACIPCGPGS---KNNQDHSVYCSDCFEYHEKENQILHYDFSMLS 691  
Db 489 CTNNYFLYMGCCYSTQK----APGSFMCKTAGNTGIC-----TEAANN----- 527  
Qy 692 SVGSLMNGPSTSGTKYFHFFNISLCGHEGKKMALCTNNITDFVKEIVAGSDDYTN-L 750  
Db 528 -----RYFVVPGASNTDQSVLACSNPLGLTL-----GTGDTAKAY 562  
Qy 751 VGFVQCSTIIPSESKGFRAALSSQSIIADTFIGTVETTLKNINIKEDMFPVPTSQIP 810  
Db 563 VGVEGCSQCTAP-----AALSDDGM----- 582  
Qy 811 DVHFFYKSSATFTSCINGRSTAVKMRNCPKSGAGVISVPSKCPAGTCDGCTFYFLWESA 870  
Db 583 -----APAVTSCDSSK-----KPNRDGSGCV---LCSVGGCKSCV----- 615

Qy 871 EACPLCTEHDFHEIEGACKRGFOETLYVWNEPKWCIKISLPEKKLATCETVDFWLKVGA 930  
Db 616 -----MDNICGECNSGF-----SLDNGKCVSSGANRSLGASAGA 648  
Qy 931 GVGAFATAVL-----LVALTCYFW 948  
Db 649 TAGISVAVVAVVGGGLVAFLCWVF 671  
RESULT 8  
T42215  
zonadhesin - mouse  
N;Alternate names: sperm-specific membrane protein  
C;Species: Mus musculus (house mouse)  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C;Accession: T42215  
R;Gao, Z.; Garbers, D.L.  
J. Biol. Chem. 273, 3415-3421, 1998  
A;Title: Species diversity in the structure of zonadhesin, a sperm-specific membrane  
A;Reference number: Z22080; MUID:98123114; PMID:9452463  
A;Accession: T42215  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-5376 <GAO>  
A;Cross-references: EMBL:U97068; NID:g3327420; PID:g3327421; PIDN:AAC26680.1  
C;Genetics:  
A;Gene: zan  
A;Map position: 5  
C;Function:  
A;Description: functions in multiple cell adhesion processes  
A;Note: found exclusively on the apical region of the sperm head  
C;Keywords: cell adhesion

Query Match 3.4%; Score 194; DB 2; Length 5376;  
Best Local Similarity 20.5%; Pred. No. 0.00048;  
Matches 144; Conservative 63; Mismatches 204; Indels 290; Gaps 38;

Qy 75 SSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVCSKCGEGTYSLSGI 134  
Db 3256 SRGCTQSCCTCPAGAIHCR-----NFKCPSGTCKNGDNGSSNCTEITLQCPNTS 3304  
Qy 135 KFEWDELPAFNSNIATFMDTVVGPDSRDPGCGNNSSWIIPRGVIESNRDDCTVSLIYAV 194  
Db 3305 QFT--DCLP-----SCVPSCSNRCEVTSVPSSCREG----- 3336  
Qy 195 HLKSGYVFEYQYVDNNIFFEFFFIONDQCQEMDTTDDKWKVLTNDNGEWSHVMKSGT 254  
Db 3337 -LNHGFVFE-----DKCVPRTQCGCKDARGAIIIPAG-- 3368  
Qy 255 NILYWRRTGILMGSKAVKPVLVKNITIEGVAYTSECFPCPKPGTFSNKPGSFNCQ--VCPR 312  
Db 3369 -----KTWTSKGCTQSCACV-----EGNIOQCQNFQCPP 3396  
Qy 313 NTYSEKAKECIRCKDDSQFSGSSSECTE--RPPCTTKDYFQIHTPC-----DEEG 360  
Db 3397 ETY-----CKDNSE--GSSTCTKITLQCPAHTQ-----YTSCLPSCLPSCLDPEG 3439  
Qy 361 KTOIMYKWIEPKI---CRE-----DLTDAT-RLPPSGEKKDCPP 395  
Db 3440 ----LCKDISPKVPSTCKEGCVQSGYVLNSDKCVLRAECDCKDAQGALIPAGKTWTSPG 3495  
Qy 396 CNPG-FYNNSSSCH--PCPPGTF----SDGTKECR---PCPAGTEPALGFYKWNVL 444  
Db 3496 CTQSCACMGGAQVQCSQCPPTGTYCKDNEDGNSNCAKITLQCPAHS-----LFTNCL 3547  
Qy 445 PGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTGAT 504  
Db 3548 PPCL-PSCLDP-----DGL-----CKGASPKVPSTCKE 3574  
Qy 505 GSELGRITVFETLCSADCVLVFMVDINRKSTN--VVESWGGTKEKQAYTHIIFKNATFT 562  
Db 3575 G-----CICQSGYVL-----SNNKCLLRNRCGCKD--AHGALIPEDK--- 3609











Db 1006 NCDROTGACLKCLYQTTGDHCELCCKDGGFFGDALQQNCQQCECDFLGTNNTIAHCDRFTGQ 1065  
QY 806 TSQIPDV-----HFFYKSSATTATSCINGRSTAVKMRCNPTKSGAGVISVPSKCP 854  
Db 1066 CPCLPNVQGVRCDOCAENHHKIASGEGCESCNDPIGALHEOCN---SYTG---QCQCK 1118  
QY 855 AG-----TCDGCTFYFLWESAEACPLCTEHDFHEIEGACKRGFQETL-----YVWNEPK 903  
Db 1119 PGFGGRACNQCQAHYWGPNPNEKQCPCECDQFGAADFQCDRETGNCVCHEGIGGYKCNE-- 1176  
QY 904 WCIKGISLPEKKLATC-ETVDFW-LKVAGVGGAFTAVLL-----VALTCYFWKK 950  
Db 1177 -CARGYIQGFPHCSPCGECFNNWDLILSALEDATTATILRAKEIKQVGATGAYTSEFSEL 1235  
QY 951 NQKLEY-----KYSKLVMTNKSKECELPAAADSCAIMEGEDNEEVVYSNKQS 997  
Db 1236 DKKLQHIRNLLQNTSVSLVDIEKLDYETQSLRDQLQASHG-RLSETEQNLDIY--NLSLS 1292  
QY 998 LLG-KLKSLATKEKEDHFESVQ 1018  
Db 1293 LSGVELESL-----QNH SRLVQ 1309  
  
RESULT 13  
T23433  
hypothetical protein K08C7.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: T23433  
R:Berk, M.  
submitted to the EMBL Data Library, March 1996  
A:Reference number: Z19740  
A:Accession: T23433  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-3672 <WIL>  
A:Cross-references: EMBL:Z70286; PIDN:CAA94293.1; GSPDB:GN00022; CESP:K08C7.3  
A:Experimental source: clone K08C7  
C:Genetics:  
A:Gene: CESP:K08C7.3  
A:Map position: 4  
A:Introns: 66/1; 284/3; 563/1; 1187/3; 1248/3; 1300/1; 1460/1; 1623/3; 2361/3; 2988/3; 3  
C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like  
  
Query Match 3.1%; Score 178.5; DB 2; Length 3672;  
Best Local Similarity 22.6%; Pred. No. 0.0033;  
Matches 116; Conservative 54; Mismatches 203; Indels 141; Gaps 30;  
  
QY 279 ITIEGVAYTSECFPCKP--GTFSNKPGEF--NCQVCPRNTYSEKGAKECIRCKDDSQFSG 334  
Db 1413 LSCDCVAQGSSEFQCEYGGQCKKPGVIGRRRCERCAPGYN---FPECIKC----QCNA 1465  
QY 335 SSECTERP-PCYTKDYFQIHTPCDEEGKTQIMYKWIPEKI-----CREDLTDAIRL--- 384  
Db 1466 GQQCDERTGQCFPPHVEGQT-CDRCVSNAGFY--DPLIGCQKCGCHPQSGEGNLVCD 1521  
QY 385 PPSGE-----KKDCPPCPNPGFYNNGSSSCHPCP---PGT---FSDGTK----- 421  
Db 1522 PESGQCLCRESMGGRQCDRLAGFY--GFPHCYGCSNCRAGTTEEICDATNAQCKCKENV 1579  
QY 422 ---ECRPCAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAG 478  
Db 1580 YGGRCEACKAGT-----FDLSAENPL-GCVNCFCFGVTDSCRSSMYPVTIMSDMSSFLT 1633  
QY 479 GSDN-----DYLILNLHIPG-----FKPPTSMGTATGSELGRITVFETL-----C 519  
Db 1634 TDDNGMVDNKDDTVIYTSEETSPNSVYFNVPDIEKKDYTTSYGLKLTFLKLTSTVPRGGRKSM 1693  
QY 520 SADCVLYFMVDINRKSTNV-VESWGG---TKEKQAYTHIIFKNATFTTWFQRTNQGD 575  
Db 1694 NAD-----ADVRLTGANMTIEYWASEQPTNPEEQFT-VKCKLVPENFLTAEGKTVTREE 1746  
  
Query Match 3.1%; Score 178.5; DB 2; Length 3672;  
Best Local Similarity 22.6%; Pred. No. 0.0033;  
Matches 116; Conservative 54; Mismatches 203; Indels 141; Gaps 30;  
  
QY 279 ITIEGVAYTSECFPCKP--GTFSNKPGEF--NCQVCPRNTYSEKGAKECIRCKDDSQFSG 334  
Db 1413 LSCDCVAQGSSEFQCEYGGQCKKPGVIGRRRCERCAPGYN---FPECIKC----QCNA 1465  
QY 335 SSECTERP-PCYTKDYFQIHTPCDEEGKTQIMYKWIPEKI-----CREDLTDAIRL--- 384  
Db 1466 GQQCDERTGQCFPPHVEGQT-CDRCVSNAGFY--DPLIGCQKCGCHPQSGEGNLVCD 1521  
QY 385 PPSGE-----KKDCPPCPNPGFYNNGSSSCHPCP---PGT---FSDGTK----- 421  
Db 1522 PESGQCLCRESMGGRQCDRLAGFY--GFPHCYGCSNCRAGTTEEICDATNAQCKCKENV 1579  
QY 422 ---ECRPCAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAG 478  
Db 1580 YGGRCEACKAGT-----FDLSAENPL-GCVNCFCFGVTDSCRSSMYPVTIMSDMSSFLT 1633  
QY 479 GSDN-----DYLILNLHIPG-----FKPPTSMGTATGSELGRITVFETL-----C 519  
Db 1634 TDDNGMVDNKDDTVIYTSEETSPNSVYFNVPDIEKKDYTTSYGLKLTFLKLTSTVPRGGRKSM 1693  
QY 520 SADCVLYFMVDINRKSTNV-VESWGG---TKEKQAYTHIIFKNATFTTWFQRTNQGD 575  
Db 1694 NAD-----ADVRLTGANMTIEYWASEQPTNPEEQFT-VKCKLVPENFLTAEGKTVTREE 1746

QY 576 NRRFINDMVKI-----YSITATNAVDGV--ASSCRACALGSEQSGS 614  
Db 1747 LMKVLHSLQNITLKASYFDHPKPTSTLYEFGLEISEPNGVDSVIKASSVEQCQC PAPYTG 1806  
QY 615 SCVPCPPGHYIEKETNQCKECPDPPTYLSIHQVYGKE---ACIPC---GPGSKNNQDHSVC 668  
Db 1807 SCQLCASGY-----HRVQSGSFLGACVPCECNGHSATCDPDTGIC 1846  
  
QY 669 YSDCFYHEKENQILHYDFSNLSSVGSGLMNGPSF 702  
Db 1847 -TDC-----EHTNGDHCFECNEGHYGNATNGSPY 1875  
  
RESULT 14  
T37316  
probable laminin alpha chain - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Jan-2000  
C:Accession: T37316  
R:Joh, K.; Zhu, K.; Hedgecock, E.M.; Inoue, T.; Horii, K.  
submitted to the EMBL Data Library, August 1998  
A:Description: Laminin alpha chain gene in the nematode C. elegans.  
A:Reference number: Z21681  
A:Accession: T37316  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-3704 <JOH>  
A:Cross-references: EMBL:AB016806; PIDN:BAA32347.1  
A:Experimental source: strain N2  
C:Genetics:  
A:Gene: epi-1  
A:Map position: IV  
A:Introns: 66/1; 284/3; 563/1; 1187/3; 1248/3; 1300/1; 1460/1; 1623/3; 2361/3; 2988/3  
C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like  
  
Query Match 3.1%; Score 178.5; DB 2; Length 3704;  
Best Local Similarity 22.6%; Pred. No. 0.0033;  
Matches 116; Conservative 54; Mismatches 203; Indels 141; Gaps 30;  
  
QY 279 ITIEGVAYTSECFPCKP--GTFSNKPGEF--NCQVCPRNTYSEKGAKECIRCKDDSQFSG 334  
Db 1413 LSCDCVAQGSSEFQCEYGGQCKKPGVIGRRRCERCAPGYN---FPECIKC----QCNA 1465  
QY 335 SSECTERP-PCYTKDYFQIHTPCDEEGKTQIMYKWIPEKI-----CREDLTDAIRL--- 384  
Db 1466 GQQCDERTGQCFPPHVEGQT-CDRCVSNAGFY--DPLIGCQKCGCHPQSGEGNLVCD 1521  
QY 385 PPSGE-----KKDCPPCPNPGFYNNGSSSCHPCP---PGT---FSDGTK----- 421  
Db 1522 PESGQCLCRESMGGRQCDRLAGFY--GFPHCYGCSNCRAGTTEEICDATNAQCKCKENV 1579  
QY 422 ---ECRPCAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAG 478  
Db 1580 YGGRCEACKAGT-----FDLSAENPL-GCVNCFCFGVTDSCRSSMYPVTIMSDMSSFLT 1633  
QY 479 GSDN-----DYLILNLHIPG-----FKPPTSMGTATGSELGRITVFETL-----C 519  
Db 1634 TDDNGMVDNKDDTVIYTSEETSPNSVYFNVPDIEKKDYTTSYGLKLTFLKLTSTVPRGGRKSM 1693  
QY 520 SADCVLYFMVDINRKSTNV-VESWGG---TKEKQAYTHIIFKNATFTTWFQRTNQGD 575  
Db 1694 NAD-----ADVRLTGANMTIEYWASEQPTNPEEQFT-VKCKLVPENFLTAEGKTVTREE 1746  
  
QY 576 NRRFINDMVKI-----YSITATNAVDGV--ASSCRACALGSEQSGS 614  
Db 1747 LMKVLHSLQNITLKASYFDHPKPTSTLYEFGLEISEPNGVDSVIKASSVEQCQC PAPYTG 1806  
QY 615 SCVPCPPGHYIEKETNQCKECPDPPTYLSIHQVYGKE---ACIPC---GPGSKNNQDHSVC 668  
Db 1807 SCQLCASGY-----HRVQSGSFLGACVPCECNGHSATCDPDTGIC 1846  
  
QY 669 YSDCFYHEKENQILHYDFSNLSSVGSGLMNGPSF 702  
Db 1847 -TDC-----EHTNGDHCFECNEGHYGNATNGSPY 1875



Db 1847 -TDC-----EHTNGDHCFCNEGHYGNATNGSPY i875

RESULT 15

MMMSB2

laminin gamma-1 chain precursor - mouse

N:Alternate names: laminin chain B2

C:Species: Mus musculus (house mouse)

C>Date: 28-Feb-1986 #sequence\_revision 30-Jun-1991 #text\_change 10-Dec-1999

C:Accession: A28469; A27729; A28082; S02680; S05327; S02037; A02870; S13544; S14552

R:Sasaki, M.; Yamada, Y.

J. Biol. Chem. 262, 17111-17117, 1987

A:Title: The laminin B2 chain has a multidomain structure homologous to the B1 chain.

A:Reference number: A28469; MUID:88059118; PMID:3680290

A:Accession: A28469

A:Molecule type: mRNA

A:Residues: 1-1607 <SAS>

A:Cross-references: EMBL:J03484; NID:g198694; PIDN:AAA39405.1; PID:g293688

R:Durkin, M.E.; Bartos, B.B.; Liu, S.H.; Phillips, S.L.; Chung, A.E.

Biochemistry 27, 5198-5204, 1988

A:Title: Primary structure of the mouse laminin B2 chain and comparison with laminin B1.

A:Reference number: A27729; MUID:89000737; PMID:3167041

A:Accession: A27729

A:Molecule type: mRNA

A:Residues: 1-263, 'D', 265-336, 'C', 338-446, 'PS', 449-661, 'S', 663-885, 887-1155, 1157-1433, 'A'

A:Cross-references: EMBL:J02930; NID:g198702; PIDN:AAA39408.1; PID:g293691

A>Note: the authors translated the codon TAT for residue 544 as Asp and GCG for residue

R:Ogawa, K.; Burbelo, P.D.; Sasaki, M.; Yamada, Y.

J. Biol. Chem. 263, 8384-8389, 1988

A:Title: The laminin B2 chain promoter contains unique repeat sequences and is active in

A:Reference number: A28082; MUID:88228071; PMID:2836421

A:Accession: A28082

A:Molecule type: DNA

A:Residues: 1-215, 'A', 217-239 <OGA>

A:Cross-references: EMBL:J03749; NID:g198704; PIDN:AAA39409.1; PID:g554184

R:Fujiwara, S.; Shinkai, H.; Deutzmann, R.; Paulsson, M.; Timpl, R.

Biochem. J. 252, 453-461, 1988

A:Title: Structure and distribution of N-linked oligosaccharide chains on various domain

A:Reference number: S02678; MUID:88326259; PMID:2458101

A:Accession: S02680

A:Molecule type: protein

A:Residues: 227-238 <FUJ>

R:Hartl, L.; Oberbaeumer, I.; Deutzmann, R.

Eur. J. Biochem. 173, 629-635, 1988

A:Title: The N terminus of laminin A chain is homologous to the B chains.

A:Reference number: S00624; MUID:88225080; PMID:3267223

A:Accession: S05327

A:Molecule type: protein

A:Residues: 227-238, 387-393, 'F', 395-405, 881-912, 1022-1034 <HAR>

R:Deutzmann, R.; Huber, J.; Schmetz, K.A.; Oberbaeumer, I.; Hartl, L.

Eur. J. Biochem. 177, 35-45, 1988

A:Title: Structural study of long arm fragments of laminin. Evidence for repetitive C-te

A:Reference number: S01790; MUID:89030693; PMID:3181157

A:Accession: S02037

A:Molecule type: protein

A:Residues: 1362-1377, 'X', 1379-1392, 'X', 1394-1406 <DEU>

R:Barlow, D.P.; Green, N.M.; Kurkinen, M.; Hogan, B.L.M.

EMBO J. 3, 2355-2362, 1984

A:Title: Sequencing of laminin B chain cDNAs reveals C-terminal regions of coiled-coil a

A:Reference number: A02870; MUID:85051302; PMID:6209134

A:Accession: A02870

A:Molecule type: mRNA

A:Residues: 1391-1474, 'K', 1476-1575, 'N', 1577-1607 <BAR>

A:Cross-references: EMBL:X05211; NID:g52862; PIDN:CAA28838.1; PID:g817975

R:Paulsson, M.; Deutzmann, R.; Timpl, R.; Dalzoppo, D.; Odermatt, E.; Engel, J.

EMBO J. 4, 309-316, 1985

A:Title: Evidence for coiled-coil alpha-helical regions in the long arm of laminin.

A:Reference number: S13543; MUID:85257455; PMID:3848400

A:Accession: S13544

A:Molecule type: protein

A:Residues: 1506-1523, 'X', 1525 <PAU>

R:Olsen, D.; Nagayoshi, T.; Fazio, M.; Peltonen, J.; Jaakkola, S.; Sanborn, D.; Sasaki,

Lab. Invest. 60, 772-782, 1989

A:Title: Human laminin: cloning and sequence analysis of cDNAs encoding A, B1 and B2  
A:Reference number: A34961; MUID:89280632; PMID:2733383  
A:Accession: S14552

A:Molecule type: protein

A:Residues: 881-912, 1022-1034, 1364-1377, 1379-1392, 1394-1409, 1506-1525, 1593-1606 <OLS>  
C:Genetics:

A:Gene: Lamb-2

A:Map position: 1

A:Introns: 138/1; 239/3

C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type lamin  
C:Function:

A:Description: Interact with cells and with other basement membrane proteins to promo  
C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology

C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellu  
F:1-33/Domain: signal sequence #status predicted <SIG>

F:34-1607/Product: laminin gamma-1 chain #status predicted <MAT>

F:34-283/Domain: VI <DOM6>

F:284-502/Domain: V <DOM5>

F:284-337/Domain: laminin-type EGF-like homology #status atypical <LE01>

F:340-393/Domain: laminin-type EGF-like homology <LE02>

F:396-440/Domain: laminin-type EGF-like homology <LE03>

F:443-490/Domain: laminin-type EGF-like homology <LE04>

F:493-502/Domain: laminin-type EGF-like homology #status atypical <LE05>

F:503-687/Domain: IV <DOM4>

F:688-1032/Domain: III <DOM3>

F:688-719/Domain: laminin-type EGF-like homology #status atypical <LE06>

F:722-768/Domain: laminin-type EGF-like homology <LE07>

F:771-823/Domain: laminin-type EGF-like homology <LE08>

F:826-879/Domain: laminin-type EGF-like homology <LE09>

F:882-930/Domain: laminin-type EGF-like homology <LE10>

F:933-978/Domain: laminin-type EGF-like homology <LE11>

F:981-1026/Domain: laminin-type EGF-like homology <LE12>

F:1033-1607/Domain: II/I <DOM2>

F:1033-1607/Region: heptad repeats

F:38-48/Disulfide bonds: #status predicted

F:58,132,574,648,1020,1105,1159,1173,1203,1221,1239,1437/Binding site: carbohydrate (

F:1029,1032/Disulfide bonds: interchain #status predicted

F:1378,1393/Binding site: carbohydrate (Asn) (covalent) #status experimental

F:1598/Disulfide bonds: interchain (to chain B1) #status experimental

Query Match 3.1%; Score 177.5; DB 1; Length 1607;  
Best Local Similarity 19.3%; Pred. No. 0.0014;  
Matches 215; Conservative 115; Mismatches 437; Indels 347; Gaps 57;

QY	13	GWGRPAEAPR-RGRSPPWSPAWICCWALAGCQAAWAGDLPSSSSRLPPCQEKDYHFEYT	71
Db	3	GGGRAALALQPRGR-----LWPLLAVALAAGCVRAAMD-----	36
QY	72	ECDSGSRWRVAIP---NSAVD-----CSGLPDPVRGKECTFSCASGEYLEMKNQVC	120
Db	37	ECADGGRPQRCMPEFVNAAFNVTVVATNTCGTPPE---EYCVQTGVTG-----VTKSC	87
QY	121	SKCGEGTYSLGSGIKF--DEWDELPAFGFSNIATFMDTVVVGPSD-----	161
Db	88	HLCDAGQQHLQHGAFLTDYNNQADTTWWSQOTMLAGVQYPNSINLTLHLGKAFDITYVR	147
QY	162	-----SRPDGC-----NNSSWIP-----RGNYESNR-----DD-----CT---	187
Db	148	LKFHFSRPESFAIYKRTREDGPWIPYQYSGSCENTYSKANRGFIRTGDEQQALCTDEF	207
QY	188	--VSLIYAVHL-----KKSgyvFEYQYVDNNIFFEFFFIONDQCQEMDTTDDKWKVLT	238
Db	208	SDISPLTGGNVAFSTLEGRPSAYNF-----DN-----SPVQEWVTATDIRVTLN	252
QY	239	DNGEWG----SHSVMLKSGTNILYWRTTGILMG-----SKAVKPVLVKNI-TIEGV	284
Db	253	RLNTFGDEVFNPKVLKS----YYAISDFAVGGRCKNGHASECVKNEFDKLMCNCKHN	308
QY	285	AYTSECFPCPKPGTFSNKP-----GFSNCQVCPRNTYSEKGAKECIR	325
Db	309	TYGVDCCKCLP-FFNDRPWRRATAESASESLPCDNGRSQECYFDPPELYRSTGHGGHCTN	367
QY	326	CKDDSQFSGSSECTER-----PPCTTKDYFQIHPTPCDEBGTQIMYKWIEPKICR	375



Db 368 CRDNTDGAKCERENFFRLGNTTEACSPCHCSPVGSLSLTCQDSYGRCS-----CK 417  
QY 376 EDLTDAILRPPSGEKKDCPPCNPFGYNNNGSSSCHPCPPGTFSDGTKECR-----PCPAG 429  
Db 418 PGV-----MGDK--CDRCQPGFHSLTEAGCRPC--SCDLRGSTDECNVETGRCVCCKDN 466  
QY 430 TEPALGFEYK-----WWNVLPGNMK--TSCFNVGNSK-CDGMNGWEV-----AG 470  
Db 467 VE---GFNCERCKPGFFNLESSNPKGCTPCFCFGHSSSVCTNAVGYSVDISSTFQIDEDG 523  
QY 471 DHIQSGAG-----GSDND--YLILNLHIPG-FKPPPTSMGTGATGSELGR-ITFVF----- 515  
Db 524 WRVEQRDGSEASLEWSSDRQDIAVISDSYFPYFIAPVKFELGNQVLSYGQNLSFSFRVDR 583  
QY 516 -ETLCSADCVLVFMVDINRKSTNVVESWGGTKEKQAYTHIIFKNATFTFTW-----AF 567  
Db 584 RDRLSAEDLVLEGAGL-RVSVPLIAQNSYPSSETTVKYIFRLHEATDYPWRPALSPFEF 642  
QY 568 QRTNQGDNRRFINDMVKI-----YSITATNAVDBGVASSCRACALGSEQSGSSCVPCPPG 622  
Db 643 Q-----KLLNLTSIKIRGYTSERTAGYLDV-----TLQSARPG---PGVPA 682  
QY 623 HYIEKETNOCKECPDPTYLSIHQVYGKEACIPCGPGSKNNQDHSVYSDCFFY----- 675  
Db 683 TWVESC-----CPVG-----YGGQFCETCLPGYRRETPSLGPYSPCVLCTCNGHSE 729  
QY 676 -----HEKENQILHYDFSNLSSVGSMLMNGPSFTSKGTKYFHHFFNISL 717  
Db 730 TCDPETGVCDCRDNTAGPHCEKCSGDGYGDDTLGTSSDCQPCPCPGSSCAIVPKTKEV 789  
QY 718 CGH-----EGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFFVCOSTIIPSESKGFRAAL 772  
Db 790 CTHCPTGTAGKRCELCDDGYFGDPL-----GSGNPVRLCRPCQCNDNIDPN-AVGNCNRL 843  
QY 773 SSQSIILADTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSTATTSINGRSTA 832  
Db 844 TGECLKCIYNTAGFYCD-----RKEGFFGNPLAPNP-----ADCKACACNPGTVQ 891  
QY 833 VKMRCNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEHDFHEIEGAC--KR 890  
Db 892 QQSSCNPVTGQCQCLPHVSGRDCGTCDD--PGYYNLQSGQGCERCDCCHALGSTNGQCDIRT 949  
QY 891 GFQETLYVWNEPKWCIKIGISLPEKKLATCETVDF 924  
Db 950 GQCE-----CQPGIT--GQHCCERCETNHF 971

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 11:12:58 ; Search time 20.6432 Seconds  
(without alignments)  
2063.447 Million cell updates/sec

Title: US-10-073-333A-4  
Perfect score: 5681  
Sequence: 1 MLFRARGPVGRGWGRPAEA.....KEKEDHFESVLKTSRSPNI 1027

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues 112892

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	248.5	4.4	1696	1	PCK5_BRACL	Q9nj15 branchiosto
2	242.5	4.3	1877	1	PCK5_MOUSE	Q04592 mus musculu
3	223.5	3.9	687	1	VS41_GIALA	P92127 giardia lam
4	210	3.7	713	1	TSA4_GIALA	P21849 giardia lam
5	200	3.5	3718	1	LMA5_MOUSE	Q61001 mus musculu
6	194	3.4	5376	1	ZAN_MOUSE	O88799 mus musculu
7	184	3.2	1790	1	LMB1_DROME	P11046 drosophila
8	182	3.2	667	1	TS11_GIALA	Q03185 giardia lam
9	178.5	3.1	3672	1	LML2_CAEEL	Q21313 caenorhabdi
10	177.5	3.1	1607	1	LMG1_MOUSE	P02468 mus musculu
11	177.5	3.1	2907	1	FBN2_MOUSE	Q61555 mus musculu
12	173.5	3.1	1581	1	LMG3_MOUSE	Q9r0b6 mus musculu
13	173.5	3.1	2871	1	FBN1_MOUSE	Q61554 mus musculu
14	173	3.0	1786	1	LMB1_MOUSE	P02469 mus musculu
15	170.5	3.0	2437	1	NTC1_BRARE	P46530 brachydanio
16	170	3.0	1557	1	LML1_CAEEL	Q18823 caenorhabdi
17	170	3.0	1786	1	LMB1_HUMAN	P07942 homo sapien
18	170	3.0	2491	1	MPRI_HUMAN	P11717 homo sapien
19	169	3.0	1639	1	LMG1_DROME	P15215 drosophila
20	167.5	2.9	1587	1	LMG3_HUMAN	Q9y6n6 homo sapien
21	166	2.9	2871	1	FBN1_HUMAN	P35555 homo sapien
22	165	2.9	3110	1	LMA2_HUMAN	P24043 homo sapien
23	164	2.9	2871	1	FBN1_BOVIN	P98133 bos taurus
24	164	2.9	2871	1	FBN1_PIG	Q9tv36 sus scrofa
25	164	2.9	3084	1	LMA1_MOUSE	P19137 mus musculu
26	161	2.8	3712	1	LMA1_DROME	Q00174 drosophila
27	160.5	2.8	1609	1	LMG1_HUMAN	P11047 homo sapien
28	160.5	2.8	2911	1	FBN2_HUMAN	P35556 homo sapien
29	158	2.8	3695	1	LMA5_HUMAN	O15230 homo sapien
30	157.5	2.8	1219	1	JAG1_RAT	Q63722 rattus norv
31	157	2.8	2531	1	NTC1_MOUSE	Q01705 mus musculu
32	156.5	2.8	3106	1	LMA2_MOUSE	Q60675 mus musculu
33	155.5	2.7	2003	1	NTC4_HUMAN	Q99466 homo sapien

34	155	2.7	2499	1	MPRI_BOVIN	P08169 bos taurus
35	154.5	2.7	2482	1	VWF_PIG	Q28833 sus scrofa
36	153.5	2.7	1680	1	FUR2_DROME	P30432 drosophila
37	153	2.7	1959	1	AGRI_RAT	P25304 rattus norv
38	153	2.7	2318	1	NTC3_MOUSE	Q61982 mus musculu
39	152	2.7	1218	1	JAG1_MOUSE	Q9qxx0 mus musculu
40	152	2.7	3707	1	PGBM_MOUSE	Q05793 mus musculu
41	151.5	2.7	2703	1	NOTC_DROME	P07207 drosophila
42	151	2.7	1231	1	CFAH_HUMAN	P08603 homo sapien
43	151	2.7	2470	1	NTC2_MOUSE	O35516 mus musculu
44	149.5	2.6	2524	1	NOTC_XENLA	P21783 xenopus lae
45	149	2.6	2321	1	NTC3_HUMAN	Q9um47 homo sapien

ALIGNMENTS

RESULT 1  
PCK5\_BRACL  
ID PCK5\_BRACL STANDARD; PRT; 1696 AA.  
AC Q9NJ15; Q9NJ16; Q9NJ14;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)  
DE (Proprotein convertase PC6-like) (aPC6).  
GN PC6.  
OS Branchiostoma californiensis (California lancelet) (Amphioxus).  
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
OC Branchiostoma.  
OX NCBI\_TaxID=7738;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).  
RX MEDLINE=20175281; PubMed=10708868;  
RA Oliva A.A. Jr., Chan S.J., Steiner D.F.;  
RT "Evolution of the prohormone convertases: identification of a  
homologue of PC6 in the protochordate amphioxus.";  
RL Biochim. Biophys. Acta 1477:338-348(2000).  
CC -!- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY  
WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE  
OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR  
PROPROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS, WHERE XAA  
CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.  
CC -!- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.  
ISOFORM B IS A TYPE I MEMBRANE PROTEIN.  
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A, B (SHOWN HERE) AND C; ARE  
PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE  
ASSISTING THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC  
RETICULUM.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.  
CC -!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF184615; AAF26300.1; -  
EMBL; AF184616; AAF26301.1; -  
EMBL; AF184617; AAF26302.1; -  
HSSP; Q99405; IMPT.  
DR MEROPS; S08.UPB; -  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR002884; P\_domain.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF01483; P; 1.





RC TISSUE=Brain, and Intestine;  
RX MEDLINE=93224489; PubMed=8468318;  
RA Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K.,  
RA Nakayama K.;  
RT "Identification and functional expression of a new member of the  
RT mammalian kex2-like processing endoprotease family: its striking  
RT structural similarity to PACE4.";  
RL J. Biochem. 113:132-135(1993).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM PC5A).  
RC TISSUE=Adrenal cortex;  
RX MEDLINE=93342056; PubMed=8341687;  
RA Lusson J., Vieau D., Hamelin J., Day R., Chretien M., Seidah N.G.;  
RT "cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a  
RT candidate proprotein convertase expressed in endocrine and  
RT nonendocrine cells.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).  
RN [4]  
RP PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.  
RX MEDLINE=97103178; PubMed=8947550;  
RA De Bie I., Marcinkiewicz M., Malide D., Lazure C., Nakayama K.,  
RA Bendayan M., Seidah N.G.;  
RT "The isoforms of proprotein convertase PC5 are sorted to different  
RT subcellular compartments.";  
RL J. Cell Biol. 135:1261-1275(1996).  
RN [5]  
RP DEVELOPMENTAL EXPRESSION.  
RX MEDLINE=96293359; PubMed=8698813;  
RA Constam D.B., Calfon M., Robertson E.J.;  
RT "SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone  
RT morphogenetic proteins at distinct sites during embryogenesis.";  
RL J. Cell Biol. 134:181-191(1996).  
RN [6]  
RP DEVELOPMENTAL EXPRESSION.  
RX MEDLINE=97436919; PubMed=9291583;  
RA Rancourt S.L., Rancourt D.E.;  
RT "Murine subtilisin-like proteinase SPC6 is expressed during embryonic  
RT implantation, somitogenesis, and skeletal formation.";  
RL Dev. Genet. 21:75-81(1997).  
CC -!- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPROTEASE ACTIVITY  
CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE  
CC OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF. MAY BE RESPONSIBLE  
CC FOR THE MATURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED  
CC IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION  
CC OF GROWTH FACTORS.  
CC -!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR  
CC PROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS, WHERE XAA  
CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.  
CC -!- SUBCELLULAR LOCATION: PC5A IS SECRETED THROUGH THE REGULATED  
CC SECRETORY PATHWAY. PC5B IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO  
CC A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH  
CC EARLY ENDOSOMES.  
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; PC5B/LONG (SHOWN HERE)  
CC AND PC5A/SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: PC5A IS EXPRESSED IN MOST TISSUES BUT IS MOST  
CC ABUNDANT IN THE INTESTINE AND ADRENALS. PC5B IS EXPRESSED IN THE  
CC INTESTINE, ADRENALS AND LUNG BUT NOT IN THE BRAIN.  
CC -!- DEVELOPMENTAL STAGE: WEAKLY EXPRESSED THROUGHOUT THE EMBRYO,  
CC BUT MARKEDLY UPREGULATED AT DISCRETE SITES DURING DEVELOPMENT. AT  
CC E6.5, PROMINENT EXPRESSION OBSERVED IN DIFFERENTIATED DECIDUA. AT  
CC E7.5, INTENSE EXPRESSION IN EXTRAEMBRYONIC ENDODERM, AMNION AND  
CC NASCENT MESODERM. AT 8.5, ABUNDANT EXPRESSION IN SOMITES AND YOLK  
CC SAC FOLLOWED BY A CONFINATION TO DERMATOTOME COMPARTMENT. BETWEEN  
CC E9.5 AND E11.5, ABUNDANT EXPRESSION IN AER (THICKENED ECTODERMAL  
CC CELLS OF LIMB BUDS). AT E12.5, EXPRESSION IN THE LIMBS IS CONFINED  
CC TO THE CONDENSING MESENCHYM SURROUNDING THE CARTILAGE. AT THIS  
CC STAGE, STRONG EXPRESSION ALSO DETECTED IN VERTEBRAL AND FACIAL  
CC CARTILAGE PRIMORDIA AND IN THE MUSCLE OF THE TONGUE. AT E16.5,  
CC ABUNDANT EXPRESSION IN EPITHELIAL CELLS OF THE INTESTINAL VILLI.  
CC ISOFORM A IS MOST ABUNDANT AT ALL STAGES BUT SIGNIFICANT LEVELS OF  
CC ISOFORM B OCCUR AT E12.5.  
CC -!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE

CC ASSISTING THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC  
CC RETICULUM.  
CC -!- DOMAIN: AC 1 AND AC 2 (CLUSTERS OF ACIDIC AMINO ACIDS) CONTAIN  
CC SORTING INFORMATION. AC 1 DIRECTS TGN LOCALIZATION AND INTERACTS  
CC WITH THE TGN SORTING PROTEIN PACS-1.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.  
CC -!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D17583; BAA04507.1; -;  
CC EMBL; D12619; BAA02143.1; -;  
CC EMBL; L14932; AAA74636.1; -;  
CC PIR; JX0248; JX0248.  
CC PIR; A48225; A48225.  
CC HSSP; Q99405; IMPT.  
CC MEROPS; S08.076; -;  
CC MGD; MGI:97515; PCSK5.  
CC InterPro; IPR000561; EGF-like.  
CC InterPro; IPR002174; Furin-like.  
CC InterPro; IPR002884; P\_domain.  
CC InterPro; IPR000209; Peptidase\_S8.  
CC Pfam; PF00082; Peptidase\_S8; 1.  
CC Pfam; PF01483; P; PARTIAL.  
CC PRINTS; PR00723; SUBTILISIN.  
CC ProDom; PD000717; P\_domain; 1.  
CC SMART; SM00181; EGF; 3.  
CC SMART; SM00001; EGF\_like; 2.  
CC SMART; SM00261; FU; 22.  
CC PROSITE; PS00136; SUBTILASE\_ASP; 1.  
CC PROSITE; PS00137; SUBTILASE\_HIS; 1.  
CC PROSITE; PS00138; SUBTILASE\_SER; 1.  
CC Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;  
CC Cleavage on pair of basic residues; Repeat; Alternative splicing;  
CC Transmembrane.  
CC SIGNAL 1 34  
CC PROPEP 35 116  
CC CHAIN 117 1877  
CC  
CC DOMAIN 117 1768  
CC TRANSMEM 1769 1789  
CC DOMAIN 1790 1877  
CC DOMAIN 117 452  
CC DOMAIN 464 602  
CC DOMAIN 638 1753  
CC DOMAIN 1825 1844  
CC DOMAIN 1856 1877  
CC SITE 116 117  
CC SITE 521 523  
CC ACT\_SITE 173 173  
CC ACT\_SITE 214 214  
CC ACT\_SITE 388 388  
CC CARBOHYD 227 227  
CC CARBOHYD 383 383  
CC CARBOHYD 667 667  
CC CARBOHYD 754 754  
CC CARBOHYD 804 804  
CC CARBOHYD 854 854  
CC CARBOHYD 951 951  
CC CARBOHYD 1016 1016  
CC CARBOHYD 1220 1220  
CC CARBOHYD 1317 1317  
CC CARBOHYD 1523 1523  
CC CARBOHYD 1711 1711  
CC CARBOHYD 1733 1733  
CC VARSPIC 878 915  
CC  
CC PROPEPTIDE CONVERTASE SUBTILISIN/KEXIN  
CC TYPE 5.  
CC EXTRACELLULAR (POTENTIAL).  
CC POTENTIAL.  
CC CYTOPLASMIC (POTENTIAL).  
CC CATALYTIC.  
CC HOMO B.  
CC CYS-RICH MOTIF (CRM) REGION.  
CC AC 1.  
CC AC 2.  
CC CLEAVAGE (AUTO-) (BY SIMILARITY).  
CC CELL ATTACHMENT SITE (POTENTIAL).  
CC CHARGE RELAY SYSTEM (BY SIMILARITY).  
CC CHARGE RELAY SYSTEM (BY SIMILARITY).  
CC CHARGE RELAY SYSTEM (BY SIMILARITY).  
CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC GEYIDDOGHQTCASCAKCGWPTQEDCISCPVTRVLD ->  
CC ATEESWAESGFCMLVKNNLCQKRVLOQLCCCKTCTFQG











FT	DOMAIN	834	855	LAMININ EGF-LIKE 11 (INCOMPLETE).
FT	DOMAIN	856	1442	LAMININ DOMAIN IV 1 (DOMAIN IV B)
FT	DOMAIN	1443	1488	LAMININ EGF-LIKE 12.
FT	DOMAIN	1489	1532	LAMININ EGF-LIKE 13.
FT	DOMAIN	1533	1581	LAMININ EGF-LIKE 14.
FT	DOMAIN	1582	1632	LAMININ EGF-LIKE 15.
FT	DOMAIN	1633	1642	LAMININ EGF-LIKE 16 (N-TERMINAL).
FT	DOMAIN	1643	1831	LAMININ DOMAIN IV 2 (DOMAIN IV A)
FT	DOMAIN	1832	1864	LAMININ EGF-LIKE 16 (C-TERMINAL).
FT	DOMAIN	1865	1914	LAMININ EGF-LIKE 17.
FT	DOMAIN	1915	1970	LAMININ EGF-LIKE 18.
FT	DOMAIN	1971	2024	LAMININ EGF-LIKE 19.
FT	DOMAIN	2025	2071	LAMININ EGF-LIKE 20.
FT	DOMAIN	2072	2118	LAMININ EGF-LIKE 21.
FT	DOMAIN	2119	2168	LAMININ EGF-LIKE 22.
FT	DOMAIN	2169	2735	DOMAIN II AND I.
FT	DOMAIN	2736	2933	LAMININ G-LIKE 1.
FT	DOMAIN	2947	3119	LAMININ G-LIKE 2.
FT	DOMAIN	3128	3296	LAMININ G-LIKE 3.
FT	DOMAIN	3337	3511	LAMININ G-LIKE 4.
FT	DOMAIN	3518	3689	LAMININ G-LIKE 5.
FT	DOMAIN	2205	2257	COILED COIL (POTENTIAL).
FT	DOMAIN	2330	2464	COILED COIL (POTENTIAL).
FT	DOMAIN	2604	2621	COILED COIL (POTENTIAL).
FT	DOMAIN	2639	2705	COILED COIL (POTENTIAL).
FT	SITE	1723	1725	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	1839	1841	CELL ATTACHMENT SITE (POTENTIAL).
FT	DISULFID	305	314	BY SIMILARITY.
FT	DISULFID	307	327	BY SIMILARITY.
FT	DISULFID	329	338	BY SIMILARITY.
FT	DISULFID	341	361	BY SIMILARITY.
FT	DISULFID	364	373	BY SIMILARITY.
FT	DISULFID	366	398	BY SIMILARITY.
FT	DISULFID	401	410	BY SIMILARITY.
FT	DISULFID	413	431	BY SIMILARITY.
FT	DISULFID	434	445	BY SIMILARITY.
FT	DISULFID	436	452	BY SIMILARITY.
FT	DISULFID	454	463	BY SIMILARITY.
FT	DISULFID	466	476	BY SIMILARITY.
FT	DISULFID	500	512	BY SIMILARITY.
FT	DISULFID	502	521	BY SIMILARITY.
FT	DISULFID	523	532	BY SIMILARITY.
FT	DISULFID	535	544	BY SIMILARITY.
FT	DISULFID	547	559	BY SIMILARITY.
FT	DISULFID	549	566	BY SIMILARITY.
FT	DISULFID	568	577	BY SIMILARITY.
FT	DISULFID	580	590	BY SIMILARITY.
FT	DISULFID	593	605	BY SIMILARITY.
FT	DISULFID	595	611	BY SIMILARITY.
FT	DISULFID	613	622	BY SIMILARITY.
FT	DISULFID	625	635	BY SIMILARITY.
FT	DISULFID	638	650	BY SIMILARITY.
FT	DISULFID	640	656	BY SIMILARITY.
FT	DISULFID	658	667	BY SIMILARITY.
FT	DISULFID	670	680	BY SIMILARITY.
FT	DISULFID	683	695	BY SIMILARITY.
FT	DISULFID	685	702	BY SIMILARITY.
FT	DISULFID	704	713	BY SIMILARITY.
FT	DISULFID	716	726	BY SIMILARITY.
FT	DISULFID	1443	1455	BY SIMILARITY.
FT	DISULFID	1445	1462	BY SIMILARITY.
FT	DISULFID	1464	1473	BY SIMILARITY.
FT	DISULFID	1476	1486	BY SIMILARITY.
FT	DISULFID	1533	1548	BY SIMILARITY.
FT	DISULFID	1535	1555	BY SIMILARITY.
FT	DISULFID	1557	1566	BY SIMILARITY.
FT	DISULFID	1569	1579	BY SIMILARITY.
FT	DISULFID	1582	1594	BY SIMILARITY.
FT	DISULFID	1584	1601	BY SIMILARITY.
FT	DISULFID	1603	1612	BY SIMILARITY.
FT	DISULFID	1615	1630	BY SIMILARITY.
FT	DISULFID	1865	1874	BY SIMILARITY.
FT	DISULFID	1867	1881	BY SIMILARITY.

FT	DISULFID	1884	1893	BY SIMILARITY.
FT	DISULFID	1896	1912	BY SIMILARITY.
FT	DISULFID	1915	1930	BY SIMILARITY.
FT	DISULFID	1917	1939	BY SIMILARITY.
FT	DISULFID	1941	1950	BY SIMILARITY.
FT	DISULFID	1953	1968	BY SIMILARITY.
FT	DISULFID	1971	1986	BY SIMILARITY.
FT	DISULFID	1973	1993	BY SIMILARITY.
FT	DISULFID	1996	2005	BY SIMILARITY.
FT	DISULFID	2008	2022	BY SIMILARITY.
FT	DISULFID	2072	2083	BY SIMILARITY.
FT	DISULFID	2074	2090	BY SIMILARITY.
FT	DISULFID	2092	2101	BY SIMILARITY.
FT	DISULFID	2104	2116	BY SIMILARITY.
FT	DISULFID	2119	2126	BY SIMILARITY.
FT	DISULFID	2121	2133	BY SIMILARITY.
FT	DISULFID	2135	2144	BY SIMILARITY.
FT	DISULFID	2147	2166	BY SIMILARITY.
FT	DISULFID	2169	2169	INTERCHAIN (PROBABLE).
FT	DISULFID	2172	2172	INTERCHAIN (PROBABLE).
FT	CARBOHYD	100	100	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	148	148	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	248	248	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	383	383	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	457	457	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	485	485	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	905	905	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	926	926	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	964	964	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1335	1335	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1534	1534	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2021	2021	N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 3.58; Score 200; DB 1; Length 3718;

Best Local Similarity 18.5%; Pred. No. 3.8e-06;

Matches	174;	Conservative	82;	Mismatches	279;	Indels	404;	Gaps	46;
---------	------	--------------	-----	------------	------	--------	------	------	-----

QY	6	RGPVRGRG-----WGRPAEAPRRRGRSPWPSPA	ICCWALAGCQAAGDLPSS--S	54
Dd	1465	RGHVIGRDCSRCATGYWGFPNCRP-----CD	CGARLCELTGQCICP	1506
QY	55	SRPLPP---CQEKDYHFE---YTECDSSG-SRWRVAIPNSAVD--	-CSGLPDVPRGKE	102
Dd	1507	PRTVPDPCLVCQPQSFGCHPLVGCEECCNSGPGVQELDTPTCDMDSGQCRCPN-VAGR	RR	1565
QY	103	CTFSCASGEY-----LEMKNOV----CSKCGETYSLGS	132	
Dd	1566	CD-TCAPGFYGYPSRCRPCDCHEAGTMASVCDPLTGQCHCKENQVQSRCDDQRVGTFS	LDA	1624
QY	133	G-----IKFDEWDELPA GFSNIATF-----MDTVVGPSDSRP-----	164	
Dd	1625	ANPKGCTRFCFGATER--CGNSNLARHEEFVDMEGWVLLSSDRQVVPHHRPETELLHAD	1682	
QY	165	-----DGCNSSWIIPRGNYIESNRDDCTVSLLIYA VHLK-KSGYVFFEYQYVDNNIFFEFF	218	
Dd	1683	LRSVADTFSELYWQAPP SYLGDRVSSYGGLHYELHSETQRGDIFIPYESRDPDV-----	1737	
QY	219	IQNDOCQ-----EMDTTTTDKWVKL-TDNGEWGSHSVLMKSGTNILYW	R	260
Dd	1738	LOGNQMSIAFLELAYPPPQGVHRQLQLVEGNFRHLETHNPVSREELMMVLAGEQLQIR	1797	
QY	261	TTGILMGSKAVKPVLVNITIEGVAYT-----SECPCPKPGTFSN	300	
Dd	1798	A----LFSDTSSSVSLRRVVLEVASEAGRPPASNVELCMCPANYRGDSQCACAPGYRD	1853	
QY	301	KPGSF--NCQVCPRNITYSEK---GAKECIRCKDDSQFSSESCTERPPCTTKDYFIHTP	355	
Dd	1854	TGGLFLGRCVPCQCHGHS DRCLPGSGIGVCQHNTGEDQCERC--RPGFVSDPSNPASP	1911	
QY	356	CD-----EEGKTQIMYKWIEPKICREDLTDAIRLPSPGEKKDCPPC	396	
Dd	1912	CVSCPCPLA VPSNNFNADGCVLRNGRTQC-----LCR-----PGYAGASCERC	1953	

```
QY 397 NPGFYNG---SSSCHPC-----PPGTFSDG---TKCRPC-----PAGTEPALGFE 437
Db 1954 APGFFGNPLVLGSSQPCDCSGNGDPNMFSDCDPLTGACRGCLRHTTGPCHCERCAPGF- 2012
QY 438 YKWN-VLPGN-MKTSFCNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFK 495
Db 2013 --YGNALLPGNCTRCDCSPCGTETCDPSG----- 2040
QY 496 PPTSMGTGATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVESWGGTKKQAYTHII 555
Db 2041 -----RCLCKAGVT-----GQRCDRCLEGYFGFEQCC----- 2067
QY 556 FKNATFTFTWAFQRTNQDNRREFINDVMKIYSITATNAVGVASSCRACALGSEQSGSS 615
Db 2068 -----GCRPCACGPAAGKGE 2082
QY 616 CVP-----CPPGHY--IEKETNOCK-----ECPP----- 637
Db 2083 CHPQSGQCHCQPGTTGPOCLECAPGYWGLPEKGCRCRCQPRGHCDPHTGHCTCPPLSGE 2142
QY 638 --DTYLSIHQVYKKEACIPCGPGSKNNQ----DHSVYSDCFFYHEKENQILHYDFSMLS 691
Db 2143 RCDTCSQQHV-----PVPKPGGGHICEVCDHCV-----VLLLD--DLE 2181
QY 692 SVGSL-----MNGPSTSTSGTKYFHFENISLCGHEGK 723
Db 2182 RAGALLPAIREQLQGINASSAAWARLHRLNASIADLQSK 2220
RESULT 6
ZAN_MOUSE STANDARD; PRT; 5376 AA.
AC O88799; O08647;
DT 16-Oct-2001 (Rel. 40, Created)
DT 16-Oct-2001 (Rel. 40, Last sequence update)
DT 16-Oct-2001 (Rel. 40, Last annotation update)
DE Zonadhesin precursor.
GN ZAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=98123114; PubMed=9452463;
RA Gao Z., Garbers D.L.;
RT "Species diversity in the structure of zonadhesin, a sperm-specific
RT membrane protein containing multiple cell adhesion molecule-like
RT domains.";
RL J. Biol. Chem. 273:3415-3421(1998).
RN [2]
RP SEQUENCE OF 4864-5376 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=97271566; PubMed=9126492;
RA Gao Z., Harumi T., Garbers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
RT zonadhesin gene (ZAN).";
RL Genomics 41:119-122(1997).
CC -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD.
CC -!- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
CC -!- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATES SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC -!- DOMAIN: THE VWFD DOMAIN 2 MAY MEDIATE COVALENT
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
```

```
CC -!- SIMILARITY: CONTAINS 3 MAM DOMAINS.
CC -!- SIMILARITY: CONTAINS 25 VWFD DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U97068; AAC26680.1; -.
CC EMBL: U83190; AAC53125.1; -.
CC MGD: MGI:106656; ZAN.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR003645; FOLN.
CC InterPro; IPR000998; MAM_domain.
CC InterPro; IPR002919; TIL_Cysrich.
CC InterPro; IPR003328; Tila_Cysrich.
CC InterPro; IPR001007; VWF_C.
CC InterPro; IPR001846; VWF_D.
CC Pfam; PF00094; vwd; 4.
CC Pfam; PF00629; MAM; 3.
CC Pfam; PF01826; TIL; 25.
CC Pfam; PF02345; Tila; 25.
CC SMART; SM00181; EGF; 2.
CC SMART; SM00274; FOLN; 11.
CC SMART; SM00137; MAM; 2.
CC SMART; SM00214; VWC; 17.
CC SMART; SM00216; VWD; 4.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01186; EGF_2; 18.
CC PROSITE; PS00740; MAM_1; FALSE_NEG.
CC PROSITE; PS50060; MAM_2; 3.
KW Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
KW Repeat.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 5376 ZONADHESIN.
FT DOMAIN 18 5310 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 5311 5337 POTENTIAL.
FT DOMAIN 5338 5376 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 45 210 MAM 1.
FT DOMAIN 215 374 MAM 2.
FT DOMAIN 377 542 MAM 3.
FT DOMAIN 547 1170 80 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
(MUCIN-LIKE DOMAIN).
FT DOMAIN 1171 1280 VWFD 1 (PARTIAL).
FT DOMAIN 1281 1669 VWFD 2.
FT DOMAIN 1670 2056 VWFD 3.
FT DOMAIN 2057 2459 VWFD 4.
FT DOMAIN 2460 2579 VWFD 5 (PARTIAL).
FT DOMAIN 2580 2699 VWFD 6 (PARTIAL).
FT DOMAIN 2700 2819 VWFD 7 (PARTIAL).
FT DOMAIN 2820 2939 VWFD 8 (PARTIAL).
FT DOMAIN 2940 3059 VWFD 9 (PARTIAL).
FT DOMAIN 3060 3179 VWFD 10 (PARTIAL).
FT DOMAIN 3180 3299 VWFD 11 (PARTIAL).
FT DOMAIN 3300 3416 VWFD 12 (PARTIAL).
FT DOMAIN 3417 3536 VWFD 13 (PARTIAL).
FT DOMAIN 3537 3656 VWFD 14 (PARTIAL).
FT DOMAIN 3657 3776 VWFD 15 (PARTIAL).
FT DOMAIN 3777 3892 VWFD 16 (PARTIAL).
FT DOMAIN 3893 4928 VWFD 17 (PARTIAL).
FT DOMAIN 4029 4148 VWFD 18 (PARTIAL).
FT DOMAIN 4149 4263 VWFD 19 (PARTIAL).
FT DOMAIN 4264 4283 VWFD 20 (PARTIAL).
FT DOMAIN 4384 4503 VWFD 21 (PARTIAL).
FT DOMAIN 4504 4623 VWFD 22 (PARTIAL).
FT DOMAIN 4624 4743 VWFD 23 (PARTIAL).
FT DOMAIN 4744 4863 VWFD 24 (PARTIAL).
FT DOMAIN 4864 5261 VWFD 25.
FT DOMAIN 5259 5295 EGF-LIKE.
```

FT	DISULFID	5263	5274	BY SIMILARITY.	
FT	DISULFID	5268	5283	BY SIMILARITY.	
FT	DISULFID	5285	5294	BY SIMILARITY.	
FT	CARBOHYD	339	339	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	499	499	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	1216	1216	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	1239	1239	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	1314	1314	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	1814	1814	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	1908	1908	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	1933	1933	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	2028	2028	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	2111	2111	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	2142	2142	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	2332	2332	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	2533	2533	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	2575	2575	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	2692	2692	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	2812	2812	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	3052	3052	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	3065	3065	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	3144	3144	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	3172	3172	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	3288	3288	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	3292	3292	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	3782	3782	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	4005	4005	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	4136	4136	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	4243	4243	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	4254	4254	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	4335	4335	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	4376	4376	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	4586	4586	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	5136	5136	N-LINKED (GLCNAC.	.) (POTENTIAL).
FT	CARBOHYD	5252	5252	N-LINKED (GLCNAC.	.) (POTENTIAL).
SQ	SEQUENCE	5376	AA; 579908 MW; 0E44DB77DF2A2620 CRC64;		
Query Match 3.4%; Score 194; DB 1; Length 5376;					
Best Local Similarity 20.5%; Pred. No. 1.7e-05;					
Matches 144; Conservative 63; Mismatches 204; Indels 290; Gaps 38;					
QY	75	SSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVCSKCGEGTYSLGSGI	134		
Db	3256	SRGCTQSTCTCPAGAIHCR-----NFKCPSGTGYCKNGDNGSSNCTEITLQCPNNS	3304		
QY	135	KFDEWDELPAFGFSNIATFMDTVVGPDSRDPDGCNNSSWIPRGNYIESNRDDCTVSLIYAV	194		
Db	3305	QFT--DCLP-----SCVPSCSNRCEVTPSPVSSCREG-----	3336		
QY	195	HLKKSQYVFFEYQYVDNNIFFEPIQNDQCQEMDTTDDKWVKLTNDGEWGSVMLKSGT	254		
Db	3337	-LCNHGEVFSE-----DKCVPRTQCGCKDARGAIIIPAG-	3368		
QY	255	NILYWRTTGILMGSKAVKPVLVKNITIEGVAYTSECFFCKPGTFSNKPGSFNCQ--VCPR	312		
Db	3369	-----KTWTSKGCTQSCACV-----EGNIQCQNFQCPP	3396		
QY	313	NTYSEKGAKECIRCKDDSQFSGSSECTE---RPPCTTKDYFQIHTPC-----DEEG	360		
Db	3397	ETY-----CKDNSE--GSSTCTKITLQCPAHTQ-----YTSLCLPSCLPSCLDPEG	3439		
QY	361	KTQIMYKWIIEPKI---CRE-----DLTDAI-RLPPSGEKKDCPP	395		
Db	3440	-----LCKDISPKVPSTCKEGVCQSGYVLNSDKCVLRAECDCKDAQGALIPAGKTWTSFG	3495		
QY	396	CNPG-FYNNGSSSCH--PCPPGTF---SDGTKECR-----PCPAGTEPALGFEYKWNVL	444		
Db	3496	CTQSCACMGGAVQCQSQCPGTGYCKDNEDGNSNCAKITLQCPAHS-----LFTNCL	3547		
QY	445	PGNMKTSFCFNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTMTGAT	504		
Db	3548	PPCL-PSCLDP-----DGL-----CKGASPKVPSTCKE	3574		

QY	505	GSELGRITVFETLCSADCVLVFMVDINRKSTN--VVESWGGTKEKQAYTHIFKNATFT	562		
Db	3575	G-----CICQSGYVL-----SNKCLLRNRCGCKD--AHGALIPEDK---	3609		
QY	563	FTWAFQRTNQGDNRRFINDVMVKIYSITATNAVGDGVASSCRACALSGSEQSGSCVPCPPG	622		
Db	3610	-TWVSR-----GCTQSC-VCTGSGIQCLSS--QCPPG	3637		
QY	623	HYI---EKETNQCKECPDPTYLSIHQVYKGEACIPCGPGSKNNQDH-----SVCYS	670		
Db	3638	AYCKDNEGSSNCARIPPCPANSHYT---DCFPPCPPSCSDPEGHCEASGPRVLSTCRE	3694		
QY	671	DCF----FYHEKENQILHYDFSNLSSVGSMLM-NGPSFTSKG	706		
Db	3695	GCLCNPGEVLDRDKCVPRECGCKDAQGALIPSGKWTSPG	3735		
RESULT 7					
LMB1_DROME					
ID	LMB1_DROME	STANDARD;	PRT;	1790	AA.
AC	P11046; Q26328; Q9X2T4; Q9VLW6;				
DT	01-JUL-1989 (Rel. 11, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Laminin beta-1 chain precursor (Laminin B1 chain).				
GN	LAMB1 OR LAMB1 OR CG7123.				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;				
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;				
OC	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Canton-S;				
RX	MEDLINE=94000382; PubMed=8397815;				
RA	Gow C.-H., Chang H.-Y., Lih C.-J., Chang T.-W., Hui C.-F.;				
RT	"Analysis of the Drosophila gene for the laminin B1 chain."				
RL	DNA Cell Biol. 12:573-587(1993).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=88210471; PubMed=3365769;				
RA	Montell D.J., Goodman C.S.;				
RT	"Drosophila substrate adhesion molecule: sequence of laminin B1 chain				
RT	reveals domains of homology with mouse."				
RL	Cell 53:463-473(1988).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Berkeley;				
RX	MEDLINE=20196006; PubMed=10731132;				
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,				
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,				
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,				
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,				
RA	Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,				
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,				
RA	Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,				
RA	Balleg R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,				
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,				
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,				
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,				
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,				
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,				
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,				
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,				
RA	Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,				
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,				
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,				
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,				
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,				
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,				
RA	Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,				
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,				
RA	Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,				





Query Match 3.2%; Score 184; DB 1; Length 1790;  
Best Local Similarity 18.5%; Pred. No. 2.2e-05;  
Matches 259; Conservative 139; Mismatches 478; Indels 526; Gaps 69;

QY 39 LAGCAAAGDLPPSSSRP-----LPPCQ-EKDYHFEYTECDSSGSRWRVAIPN- 86  
Db 12 LALLSQWD-PVDSQRPQHGRDRPKYPNKFIKTHPCERSSCYPATGNLLIGREN 70

QY 87 -SAVDCSGLPDPVR-----GKECTFSCASGEYLEMKNQVCSKCGEGTYSLSGSIKFD 137  
Db 71 LTASSTCGLHSPERCILSHLQKKC-FLCDTREETKHDPYKNHRIGQIIYKTKPGTNIP 129

QY 138 EWDELPAFNSN-----IATFMDTVVGPDSRDPDGCNNSSWIPRGNIES 181  
Db 130 TWQSENGKENATIQLDEAEFHFLIITF-----TTERPAAMYIER 172

QY 182 NRDDCTVSLIYAVHLKSGYVFFEYQYVDNFIFFEQNDQCQE-----MDTTDK 233  
Db 173 SFD-----FGQTHIYRY-----FAY-----DCKESFPGVPTVLENITDV 207

QY 234 W-----VKLTDNGE-----WGSHSVLMKSGTNILYWRGTGILMGSKAV 271  
Db 208 MCTSRYSNVEPSRNGEVIFRVLPPNINVTDPYAEHVQNQLKMTNLRQTKLHKLGDNLL 267

QY 272 KPV-----VKNITIEG---VAYTSECFPCPKGTFS---NKPG----- 303  
Db 268 DSRLENEEKYYGISMVMVRGSCSYCHASQCLPLDP-AFSQADNEDGMVHGRCECTHNT 326

QY 304 -SFNCQVC-----PRNTYSEKGAKECIRCK-----DDSFSGS----- 335  
Db 327 KGMNCECEDEFNDLPWKPAFGKKTACKKCECNDHAVSCHFEAVFTASGFVSGVCDN 386

QY 336 -----SECTERPPCTTKDYFQ-----IHTPCD-----EEGKTQIMYKWIPEKI-- 373  
Db 387 CLHNTRGQHCEECMPYFYRDPEQDITSERVQPCDCDPQGSDDGICDSLNELEGAVAG 446

QY 374 ---CREDLTDAIRLPPSGEKKDCPPCNPFGYN---NGSSSCHPC---PPGTFSDG----- 419  
Db 447 ACHCKAFVTG-----RRCNQCKDGYWNLQSDNPEGCEPCTCNPLGTNNSGCVMR 496

QY 420 -----TKECRPC-----PAGTE---PALGFEYKWNVLPGNMKTSCFN 454  
Db 497 TGECKCKKYVTGKDCNOCMPETYGLSEPEGCSLCNCDAGGSYDNCVVISGQCRPHM 556

QY 455 VGNSKCDGMNGW-----EVAGDHQSGAGGS-----DN 482  
Db 557 TGRSCSQPKQNYFIPLPEVHEAEVDEICISYGANGNCSLVAETPDGSFTGIGFTRVPEN 616

QY 483 DYLIILNL-HIPGFKP-----PTS-----MTGATGSEL 508  
Db 617 SELVFTVGDIPRSMYPYDAVIRYQSTSRGDWENAFITLVRPDQVDEGGCGELAAATSET 676

QY 509 GRITF-----VFETLCSADCVLVFMVDINRK-----STNVVES----- 541  
Db 677 -RIPFSLPDRSRQVVALNEVCLEAGKVYKFRYFERKRRHVDSPATILVDSLTLIPRID 735

QY 542 ----WGGTK-----EKQAYTHIIFKNATFTFTWAFQRTNQGDN---RRFINDMVKIYSITA 591  
Db 736 VTPIFGSVLADIRKKDYEKYNCKSSLYDMNYKSDPKCQNLNDILSVFVHDGASMCNCNP 795

QY 592 TNAVDBGVASSCRA-CALGSEQSGSSCVPCPPG-----HYIEKET 629  
Db 796 TGLSKVCSNGGYCQCKPNVWGRCQDQCAPGTGFGPEGCKACDCNSIGSKDKYCDLIT 855

QY 630 NOCKECPDPTVLSIHQVYGKEACIPCGPGSKNNQDHSVCS-----DCFF 674  
Db 856 GQC-QCVPNT-----YGRE-CNQCOPGYWNFPECRCVCQCNGHAATCDPIQGTICD-- 904

QY 675 YHEKENQILHYDFSLSVGLMNG-----PSFTSK-----GTKYF 710  
Db 905 -----QDSTTGYSKDCSLDGYGNPLFGSEIGCRPCRCPETVASGLAHADGCSLD 954

QY 711 HFFNISLC---CHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFVCQSTIIPSES- 765  
Db 955 TRNNMLCHCQEGYSGSRCEICADNF-----FGNPDNGGTCCKCECSNNVDLYDTG 1005

QY 766 -----KGFRALSSQSIILADTFFIGVTVETTLK-----NINIKEDMFPVP 805  
Db 1006 NCDROTGACLKCLYQTTGDHCELCCKDGGFFGDALQONCQCECDFLGTNNTIAHCDRFTGQ 1065

QY 806 TSQIPDV-----HFFYKSSATTTCINGRSTAVKMRNCNPTKSGAGVISVPSKCP 854  
Db 1066 CPCLPNVQGVRCDOQAENHWKIASGEGCESCNCDPIGALHEQCN---SYTG-----QCQCK 1118

QY 855 AG---TCDGCTFYFLWESAEACPLCTEHDFHEIEGACKRGFQETL-----YVWNEPK 903  
Db 1119 PGFGGRACNQCOAHYWGPNNEKQPCCECDQFGAADFOCDRETGNCVCHEGIGYKCN-- 1176

QY 904 WCIKGISLPEKKLATC-ETVDFW-LKVGAGVGAFATVLL-----VALTCYFWKK 950  
Db 1177 -CARGYIGQFPHCSPGCECFNNWDLILSALEDATTATILRAKEIKQVGATGAYTSEFSEL 1235

QY 951 NQKLEY-----KYSKLVMTTNSKECELPAADSCAIMEGEDNEEEVVSNNKQS 997  
Db 1236 DKKLQHIRNLLONTSVSLVDIEKLDYETQSLRDLQASHG-RLSETEQNLDIY--NSLS 1292

QY 998 LLG-KLKSLATKEKEDHFESVQ 1018  
Db 1293 LSGVELES-----QNHSLRVQ 1309

RESULT 8  
TS11\_GIALA  
ID TS11\_GIALA STANDARD; PRT; 667 AA.  
AC Q03185;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Major surface trophozoite antigen 11 precursor.  
GN TSP11.  
OS Giardia lamblia (Giardia intestinalis).  
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.  
OX NCBI\_TaxID=5741;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Isolate AD-1;  
RX MEDLINE=93241215; PubMed=8479449;  
RA Ey P.L., Khanna K., Manning P.A., Mayrhofer G.;  
RT "A gene encoding a 69-kilodalton major surface protein of Giardia  
intestinalis trophozoites."  
RL Mol. Biochem. Parasitol. 58:247-258(1993).  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE  
CC PLASMA MEMBRANE.  
CC -!- DOMAIN: CONTAINS 27 REPEATS OF A CXXC MOTIF.  
CC -!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M95814; AAA02687.1; -  
CC PIR; A48579; A48579.  
CC HSSP; P02468; 1KLO.  
CC InterPro; IPR000561; EGF-like.  
CC InterPro; IPR002174; Furin-like.  
CC InterPro; IPR005127; Giardia\_vsp.  
CC Pfam; PF03302; VSP; 1.  
CC SMART; SM00181; EGF; 3.  
CC SMART; SM00261; FU; 5.  
KW Antigen; Repeat; Transmembrane; Signal.  
FT SIGNAL 1 17





FT DOMAIN 2084 2131 LAMININ EGF-LIKE 22.  
FT DOMAIN 2693 2884 LAMININ G-LIKE 1.  
FT DOMAIN 2896 3066 LAMININ G-LIKE 2.  
FT DOMAIN 3072 3235 LAMININ G-LIKE 3.  
FT DOMAIN 3310 3482 LAMININ G-LIKE 4.  
FT DOMAIN 3488 3669 LAMININ G-LIKE 5.  
FT DISULFID 298 307 BY SIMILARITY.  
FT DISULFID 300 320 BY SIMILARITY.  
FT DISULFID 322 331 BY SIMILARITY.  
FT DISULFID 334 354 BY SIMILARITY.  
FT DISULFID 357 366 BY SIMILARITY.  
FT DISULFID 359 391 BY SIMILARITY.  
FT DISULFID 394 403 BY SIMILARITY.  
FT DISULFID 406 424 BY SIMILARITY.  
FT DISULFID 427 438 BY SIMILARITY.  
FT DISULFID 429 445 BY SIMILARITY.  
FT DISULFID 447 456 BY SIMILARITY.  
FT DISULFID 459 469 BY SIMILARITY.  
FT DISULFID 472 484 BY SIMILARITY.  
FT DISULFID 474 491 BY SIMILARITY.  
FT DISULFID 493 502 BY SIMILARITY.  
FT DISULFID 505 516 BY SIMILARITY.  
FT DISULFID 519 531 BY SIMILARITY.  
FT DISULFID 521 538 BY SIMILARITY.  
FT DISULFID 540 549 BY SIMILARITY.  
FT DISULFID 552 561 BY SIMILARITY.  
FT DISULFID 564 576 BY SIMILARITY.  
FT DISULFID 566 583 BY SIMILARITY.  
FT DISULFID 585 594 BY SIMILARITY.  
FT DISULFID 597 607 BY SIMILARITY.  
FT DISULFID 610 622 BY SIMILARITY.  
FT DISULFID 612 629 BY SIMILARITY.  
FT DISULFID 631 640 BY SIMILARITY.  
FT DISULFID 643 653 BY SIMILARITY.  
FT DISULFID 656 668 BY SIMILARITY.  
FT DISULFID 658 674 BY SIMILARITY.  
FT DISULFID 676 685 BY SIMILARITY.  
FT DISULFID 688 698 BY SIMILARITY.  
FT DISULFID 701 715 BY SIMILARITY.  
FT DISULFID 703 724 BY SIMILARITY.  
FT DISULFID 726 735 BY SIMILARITY.  
FT DISULFID 738 753 BY SIMILARITY.  
FT DISULFID 756 770 BY SIMILARITY.  
FT DISULFID 758 777 BY SIMILARITY.  
FT DISULFID 779 788 BY SIMILARITY.  
FT DISULFID 791 806 BY SIMILARITY.  
FT DISULFID 1415 1427 BY SIMILARITY.  
FT DISULFID 1417 1434 BY SIMILARITY.  
FT DISULFID 1436 1445 BY SIMILARITY.  
FT DISULFID 1448 1458 BY SIMILARITY.  
FT DISULFID 1461 1469 BY SIMILARITY.  
FT DISULFID 1463 1476 BY SIMILARITY.  
FT DISULFID 1478 1487 BY SIMILARITY.  
FT DISULFID 1490 1503 BY SIMILARITY.  
FT DISULFID 1506 1520 BY SIMILARITY.  
FT DISULFID 1508 1527 BY SIMILARITY.  
FT DISULFID 1529 1538 BY SIMILARITY.  
FT DISULFID 1541 1551 BY SIMILARITY.  
FT DISULFID 1554 1566 BY SIMILARITY.  
FT DISULFID 1556 1573 BY SIMILARITY.  
FT DISULFID 1575 1584 BY SIMILARITY.  
FT DISULFID 1587 1602 BY SIMILARITY.  
FT DISULFID 1830 1839 BY SIMILARITY.  
FT DISULFID 1832 1846 BY SIMILARITY.  
FT DISULFID 1849 1858 BY SIMILARITY.  
FT DISULFID 1861 1877 BY SIMILARITY.  
FT DISULFID 1880 1894 BY SIMILARITY.  
FT DISULFID 1882 1905 BY SIMILARITY.  
FT DISULFID 1907 1916 BY SIMILARITY.  
FT DISULFID 1919 1934 BY SIMILARITY.  
FT DISULFID 1937 1951 BY SIMILARITY.  
FT DISULFID 1939 1958 BY SIMILARITY.  
FT DISULFID 1961 1970 BY SIMILARITY.

FT DISULFID 1973 1987 BY SIMILARITY.  
FT DISULFID 1990 2000 BY SIMILARITY.  
FT DISULFID 1992 2007 BY SIMILARITY.  
FT DISULFID 2009 2018 BY SIMILARITY.  
FT DISULFID 2021 2034 BY SIMILARITY.  
FT DISULFID 2037 2048 BY SIMILARITY.  
FT DISULFID 2039 2055 BY SIMILARITY.  
FT DISULFID 2057 2066 BY SIMILARITY.  
FT DISULFID 2069 2081 BY SIMILARITY.  
FT DISULFID 2084 2096 BY SIMILARITY.  
FT DISULFID 2086 2103 BY SIMILARITY.  
FT DISULFID 2105 2114 BY SIMILARITY.  
FT DISULFID 2117 2129 BY SIMILARITY.  
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 511 511 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 634 634 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 761 761 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1014 1014 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1341 1341 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1705 1705 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1756 1756 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1868 1868 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1944 1944 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1986 1986 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2002 2002 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2159 2159 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2207 2207 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2231 2231 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2235 2235 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2401 2401 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2421 2421 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2487 2487 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 2821 2821 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3087 3087 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3242 3242 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 3541 3541 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 3672 AA; 404223 MW; 28E262DB5FF14BFA CRC64;

Query Match 3.1%; Score 178.5; DB 1; Length 3672;  
Best Local Similarity 22.6%; Pred. No. 0.00015;  
Matches 116; Conservative 54; Mismatches 203; Indels 141; Gaps 30;

QY 279 ITIEGVAYTSECFPKP--GTFESNKPGSF--NCQVCPRNTYSEKAKECIRCKDDSQFSG 334  
: : : || || || : | ||| : : : | : : ||| : | :  
Db 1413 LSCDCVAQGSSEFQCEQYGQCKCKPGVIGRRRCERCAPGYN---FPECIKC---QCNA 1465  
QY 335 SSECTERP-PCTTKDYFQIHPTPCDEBEGTKQIMYKWIIEPKI-----CREDLTDAIRL--- 384  
: | || | : : : ||| : : : | : : | : : :  
Db 1466 GQQCDERTGQCFPPHVEGQT-CDRCVSNAGFY---DPLIGCKGCGHPQSGEGGNLVCD 1521  
QY 385 PPSGE-----KKDCPPCPNPGFYNNNGSSSCHPCP---PGT---FSDGTK----- 421  
| ||| : | | ||| | : | ||| : | : | : | :  
Db 1522 PESGQCLCRESMGGRCQDRCLAGFY--GFPHCYGCSNRRAGTTEICDATNAQCKCKENV 1579  
QY 422 ---ECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNKCDGMNGWEVAGDHIQSGAG 478  
| | ||| : : | : | : | : | : | : | : | :  
Db 1580 YGGRCEACKAGT-----FDLSAENPL-GCVNCFCFGVTDSCRSSMYPVTIMSDVMSSFLT 1633  
QY 479 GSDN-----DYLILNLHIPG-----FKPPTSMGTATGSELGRITFVFETL-----C 519  
|| | : | : | : | : | : | : | : | : | :  
Db 1634 TDDNGMVDNKDDTVIYTSEETSPNSVYFENVPIEKKDYTTSGLKLTFKLSTVPRGGRKSM 1693  
QY 520 SADCVLYFMVDINRKSTNV-VESWGG---TKEQAYTHIIFKNATFTFTWAFORTNQGD 575  
: || : : : | : : : | : : | : | : | : :  
Db 1694 NAD-----ADVRLTGANMTIEYWASEQPTNPPEEQFT-VKCKLVPENFLTAEGKTVTREE 1746  
QY 576 NRRFINDMVKI-----YSITATNAVDGV--ASSCRACALGSEQSGS 614  
: : : : | : | | | | | : | : | :

Db 1747 LMKVLHSLQNTLTKASYFDHPKSTLYEFGLEISEPNGVDSVIKASSVEQCQCPCAPYTG 1806

Qy 615 SCVCPGPHYIEKETNQCKECPDPTYLSIHQVYKGE---ACIPC---GPGSKNNQDHSVC 668

Db 1807 SCQLCASYG-----HRVQSGSFLGACVPCECNHGSATCDPDTGIC 1846

Qy 669 YSDCFYHEKENQILHYDFSNLSSVGSMLMNGPSF 702

Db 1847 -TDC-----EHNTNGDHCFECNEGHYGNATNGSPY 1875

RESULT 10

LMG1\_MOUSE

ID LMG1\_MOUSE STANDARD; PRT; 1607 AA.

AC P02468;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Laminin gamma-1 chain precursor (Laminin B2 chain).

GN LAMC1 OR LAMC-1 OR LAMB-2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RN SEQUENCE FROM N.A.

RX MEDLINE=88059118; PubMed=3680290;

RA Sasaki M., Yamada Y.;

RT "The laminin B2 chain has a multidomain structure homologous to the

RT B1 chain.";

RL J. Biol. Chem. 262:17111-17117(1987).

RN [2]

RN SEQUENCE FROM N.A.

RX MEDLINE=89000737; PubMed=3167041;

RA Durkin M.E., Bartos B.B., Liu S.-H., Phillips S.L., Chung A.E.;

RT "Primary structure of the mouse laminin B2 chain and comparison with

RT laminin B1.";

RL Biochemistry 27:5198-5204(1988).

RN [3]

RN SEQUENCE OF 1-239 FROM N.A.

RX MEDLINE=88228071; PubMed=2836421;

RA Ogawa K., Burbelo P.D., Sasaki M., Yamada Y.;

RT "The laminin B2 chain promoter contains unique repeat sequences and

RT is active in transient transfection.";

RL J. Biol. Chem. 263:8384-8389(1988).

RN [4]

RN SEQUENCE OF 1391-1607 FROM N.A.

RX MEDLINE=85051302; PubMed=6209134;

RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;

RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of

RT coiled-coil alpha-helix.";

RL EMBO J. 3:2355-2362(1984).

RN [5]

RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 771-932.

RX MEDLINE=96196434; PubMed=8648630;

RA Stetefeld J., Mayer U., Timpl R., Huber R.;

RT "Crystal structure of three consecutive laminin-type epidermal growth

RT factor-like (LE) modules of laminin gammal chain harboring the

RT nidogen binding site.";

RL J. Mol. Biol. 257:644-657(1996).

RN [6]

RP STRUCTURE BY NMR OF 824-881.

RX MEDLINE=96196435; PubMed=8648631;

RA Baumgartner R., Czisch M., Mayer U., Poeschl E., Huber R.,

RA Timpl R., Holak T.A.;

RT "Structure of the nidogen binding LE module of the laminin gammal

RT chain in solution.";

RL J. Mol. Biol. 257:658-668(1996).

CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin

CC is thought to mediate the attachment, migration, and organization

CC of cells into tissues during embryonic development by interacting

CC with other extracellular matrix components.

CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three

CC different polypeptide chains (alpha, beta, gamma), which are bound

CC to each other by disulfide bonds into a cross-shaped molecule

CC Comprising one long and three short arms with globules at each

CC end.

CC THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ),

CC LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),

CC LAMININ-6 (K-LAMININ) AND LAMININ-7 (KS-LAMININ).

CC -1- SUBCELLULAR LOCATION: Extracellular.

CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR

CC COMPONENT).

CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT

CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.

CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.

CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).

CC -1- SIMILARITY: CONTAINS 11 LAMININ EGF-LIKE DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; X05211; CAA28838.1; -

DR EMBL; J03484; AAA39405.1; -

DR EMBL; J02930; AAA39408.1; -

DR EMBL; J03749; AAA39409.1; -

DR PIR; A28469; MMMSB2.

DR PDB; 1KLO; 20-AUG-97.

DR PDB; 1TLE; 12-FEB-97.

DR MGD; MGI:99914; Lamc1.

DR InterPro; IPR004089; Chmtaxis\_transd.

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR001886; LamNT.

DR InterPro; IPR000034; Laminin\_B.

DR InterPro; IPR002049; Laminin\_EGF.

DR Pfam; PF00052; laminin\_B; 1.

DR Pfam; PF00053; laminin\_EGF; 10.

DR Pfam; PF00055; laminin\_Nterm; 1.

DR PRINTS; PRO0011; EGFLAMININ.

DR ProDom; PD002082; LamNT; 1.

DR ProDom; PD003031; Laminin\_B; 1.

DR SMART; SM00180; EGF\_Lam; 9.

DR SMART; SM00001; EGF\_like; 1.

DR SMART; SM00281; LamB; 1.

DR SMART; SM00136; LamNT; 1.

DR PROSITE; PS00022; EGF\_1; 8.

DR PROSITE; PS01186; EGF\_2; 2.

DR PROSITE; PSQ1248; LAMININ\_TYPE\_EGF; 10.

KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;

KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.

FT SIGNAL 1 33

FT CHAIN 34 1607 LAMININ GAMMA-1 CHAIN.

FT DOMAIN 34 283 LAMININ N-TERMINAL (DOMAIN VI).

FT DOMAIN 284 339 LAMININ EGF-LIKE 1.

FT DOMAIN 340 395 LAMININ EGF-LIKE 2.

FT DOMAIN 396 442 LAMININ EGF-LIKE 3.

FT DOMAIN 443 492 LAMININ EGF-LIKE 4.

FT DOMAIN 493 502 LAMININ EGF-LIKE 5 (N-TERMINAL).

FT DOMAIN 503 687 LAMININ DOMAIN IV.

FT DOMAIN 688 721 LAMININ EGF-LIKE 5 (C-TERMINAL).

FT DOMAIN 722 770 LAMININ EGF-LIKE 6.

FT DOMAIN 771 825 LAMININ EGF-LIKE 7.

FT DOMAIN 826 881 LAMININ EGF-LIKE 8 (NIDOGEN-BINDING).

FT DOMAIN 882 932 LAMININ EGF-LIKE 9.

FT DOMAIN 933 980 LAMININ EGF-LIKE 10.

FT DOMAIN 981 1028 LAMININ EGF-LIKE 11.

FT DOMAIN 1029 1607 DOMAIN II AND I.

FT DOMAIN 1034 1594 COILED COIL (POTENTIAL).

FT DISULFID 340 349 BY SIMILARITY.

FT DISULFID 342 365 BY SIMILARITY.







FT	DISULFID	677	690	BY SIMILARITY.
FT	DISULFID	765	777	BY SIMILARITY.
FT	DISULFID	772	786	BY SIMILARITY.
FT	DISULFID	788	801	BY SIMILARITY.
FT	DISULFID	807	819	BY SIMILARITY.
FT	DISULFID	814	828	BY SIMILARITY.
FT	DISULFID	830	843	BY SIMILARITY.
FT	DISULFID	849	859	BY SIMILARITY.
FT	DISULFID	854	868	BY SIMILARITY.
FT	DISULFID	870	883	BY SIMILARITY.
FT	DISULFID	952	964	BY SIMILARITY.
FT	DISULFID	959	973	BY SIMILARITY.
FT	DISULFID	975	988	BY SIMILARITY.
FT	DISULFID	1070	1082	BY SIMILARITY.
FT	DISULFID	1077	1091	BY SIMILARITY.
FT	DISULFID	1093	1106	BY SIMILARITY.
FT	DISULFID	1112	1124	BY SIMILARITY.
FT	DISULFID	1119	1133	BY SIMILARITY.
FT	DISULFID	1135	1149	BY SIMILARITY.
FT	DISULFID	1155	1167	BY SIMILARITY.
FT	DISULFID	1162	1176	BY SIMILARITY.
FT	DISULFID	1178	1191	BY SIMILARITY.
FT	DISULFID	1197	1209	BY SIMILARITY.
FT	DISULFID	1204	1218	BY SIMILARITY.
FT	DISULFID	1220	1233	BY SIMILARITY.
FT	DISULFID	1239	1250	BY SIMILARITY.
FT	DISULFID	1246	1259	BY SIMILARITY.
FT	DISULFID	1261	1274	BY SIMILARITY.
FT	DISULFID	1280	1292	BY SIMILARITY.
FT	DISULFID	1287	1301	BY SIMILARITY.
FT	DISULFID	1303	1316	BY SIMILARITY.
FT	DISULFID	1322	1334	BY SIMILARITY.
FT	DISULFID	1329	1343	BY SIMILARITY.
FT	DISULFID	1345	1358	BY SIMILARITY.
FT	DISULFID	1364	1377	BY SIMILARITY.
FT	DISULFID	1371	1386	BY SIMILARITY.
FT	DISULFID	1388	1399	BY SIMILARITY.
FT	DISULFID	1405	1418	BY SIMILARITY.
FT	DISULFID	1412	1427	BY SIMILARITY.
FT	DISULFID	1429	1440	BY SIMILARITY.
FT	DISULFID	1446	1458	BY SIMILARITY.
FT	DISULFID	1453	1467	BY SIMILARITY.
FT	DISULFID	1469	1482	BY SIMILARITY.
FT	DISULFID	1488	1499	BY SIMILARITY.
FT	DISULFID	1494	1508	BY SIMILARITY.
FT	DISULFID	1510	1523	BY SIMILARITY.
FT	DISULFID	1529	1540	BY SIMILARITY.
FT	DISULFID	1535	1549	BY SIMILARITY.
FT	DISULFID	1551	1564	BY SIMILARITY.
FT	DISULFID	1647	1659	BY SIMILARITY.
FT	DISULFID	1654	1668	BY SIMILARITY.
FT	DISULFID	1670	1683	BY SIMILARITY.
FT	DISULFID	1689	1701	BY SIMILARITY.
FT	DISULFID	1696	1710	BY SIMILARITY.
FT	DISULFID	1712	1725	BY SIMILARITY.
Query Match 3.1%; Score 177.5; DB 1; Length 2907;				
Best Local Similarity 20.1%; Pred. No. 0.00013;				
Matches 224; Conservative 101; Mismatches 375; Indels 417; Gaps 67;				
QY	41	GCOAAWAGDL-----PSSSRPLPPCQEKDYHFYTECDSSGSRWR--VAIPNSAVDC	91	
Db	17	GCVALWAQGTGQPPPPKTLWPQPPPPQVPRPAVAGSEG6FMGPEYRDEGAVRASVRR	76	
QY	92	SGLPDPVRGKE-C-----TFSCASGEYLEMKNQ-----VC-SKCGEGTYS-----	129	
Db	77	RGOQELRGNVCGSRFHSYCCPGWKTLPGGNQCIVPICRNSCGDGFCSRPNMCTCSSGQ	136	
QY	130	-----LGSIGKFEWDELPAGFSNIATFMDTVVGPSDSRP---DGCNN	169	
Db	137	ISPTCGRKSIQCCSVRCMNGGTCADHDHCQCQKGY-----IGTYCGQPVCENGCON	186	
QY	170	SSWIPRGNYESNRDDCTVSLIYAVHLKSGYVF----FEYQYVDNNIFFEFFIQNDQCQ	225	

Db	187	G-----GRCIGNRCACV-----YGFTGPQCDERTYRTPCFTQ--VNNQMCQ	226	
QY	226	EMDT---TTDKWVKLTDNGEGSHSVMLKSGTNILYWRTTGILMGSKAVKPVLVKNITIE	282	
Db	227	GOLTGIVCTKTLCATIGRAWGHPCEMCPA-----QPQPCRPGFIPNIRTG	272	
QY	283	GVAYTSEC--FP--CKPGTFSNKPQSFNCQVCPRNTYSEKGAKECIRCKDDSQFSGSSEC	338	
Db	273	ACQDVDECOAIPGLCOGGNCINTVGSFECR-CP-----AGHKQ-----SET	312	
QY	339	TERPPCTTKDYFQIHTPCDEEGKTQIMYKWIPEKICREDLTDAIRLPPSGEKKDCPPCNP	398	
Db	313	TQK-----C-EDIDECSVIPGVCETGDC-----	334	
QY	399	GFYNN--GSSSCHPCPPG--TFSDGTKECRPCPACTEPALGFYKWNVLPGNMKTS---	451	
Db	335	---SNTVGSYFC-LCPRGFVTSTDGSR-CIDQRACTCFGLVNGRCAQELPGRMAKAQCC	389	
QY	452	----CFNVGN--SKC-----DGMNGWEVAGDHIQ--SGAGGSDNDYLILN-	488	
Db	390	CEPGRCSIGTPEACPVRGSEYRRLCLDGLPMGIPGSSVSRPGGTGSNGNGYGPGGT	449	
QY	489	--LHIP---GFKPPTSMTGATGSELGRI---TFVFET-----LC-----	519	
Db	450	GFLPIPGDNGFSPGVGAGVGAGGQGPITGLTILNOTIDICKHHANLCLNGRCIPTVSS	509	
QY	520	-----SADCVLYFMVDINRKSTNVVESWGGTKEKQAY---THIIFKNATFTF	563	
Db	510	YRCRCNMGYKQDANGDC-----IDVDECTSNPCSNCGDCVNTPGSYCKCHA-----	555	
QY	564	TWAFQRTNQGQ-----DNRREINDMVKIYSI-----TATNAVD-GV	598	
Db	556	--GFQRTPTKQACIDIDECIQNGVLCKNGRCVNSDGSFQCICNAGFELTTDGKNCVDHDE	613	
QY	599	ASSCRACALG---SEQSGSSCVPCPPGPHYIEKETNQC--KEC-PPDTYLSIHQV--YGK	649	
Db	614	CTTNNMCLNGMCINEDSGFKCV-CKPGFILAPNGRYCTDVDECQTPGICMNGHCINNEGS	672	
QY	650	EACIPCGPGSKNNQDHSVYCY-----SDCFFYHEKENQILHYDFSLSVGSGLMNGPSF--	702	
Db	673	FRC-DCPPGLAVGVDRVCVDTHMIRSTCYGEIKKVCVRRFPFGAVTKSECCCAN-PDYGF	730	
QY	703	-----TSKGTKYFH-----FFNISLCGHEGKKMAL-----CTNNITDFTVKEIVAGSDD	746	
Db	731	GEPCQPCPAKNSAEFHGLCSSGIGITVDGRDINECALDPDICANGICE-----	778	
QY	747	YTNLVGAFVCQSTIIPSESKGFRAALSSQSIILADTFIGTVTETTLKNINIKED-----	800	
Db	779	--NLRGTYRCNC-----NSGYEPDASGRNCIDIDECI--VNRLLCDNGLCRNTPGSYS	827	
QY	801	-----MFPVPVTSQIPDVHFFYKSSSTATTSINGRST---AVKMRCNPTK--SGAGVI	847	
Db	828	CTCPPGYVLPTEETCEDV-----NECESNPCVNGACRNNLGSFHCCECPGSKLSSTGLI	882	
QY	848	SVPSKCPAGTC-----DGCTFFYF-----LWESAEACPLCTEHDFHEIEG	886	
Db	883	CIDSL--KGTCLWLNIOQNRCEVNINGATLKSECCATLGGAWGS--PCERC-----ELDA	932	
QY	887	ACKRGFOETLYVWNEPKWCIKIGISLPEKKLATCETVD	923	
Db	933	ACPRGFAR-----IKGV-----TCEDVN	950	
RESULT 12				
LMG3_MOUSE				
ID	LMG3_MOUSE	STANDARD;	PRT;	1581 AA.
AC	Q9R0B6; Q9WTW6;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Laminin gamma-3 chain precursor (Laminin 12 gamma 3).			
GN	LAMC3.			

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Albus A.M., Burgeson B., Champlaud M.-F., Koch M., Olson P.;  
RT "Mouse laminin 12 gamma 3 chain.";  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 1-1526 FROM N.A.  
RX MEDLINE=99253969; PubMed=10318827;  
RA Iivanainen A., Morita T., Tryggvason K.;  
RT "Molecular cloning and tissue-specific expression of a novel murine  
laminin gamma3 chain.";  
RL J. Biol. Chem. 274:14107-14111(1999).  
CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin  
is thought to mediate the attachment, migration, and organization  
of cells into tissues during embryonic development by interacting  
with other extracellular matrix components.  
CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three  
different polypeptide chains (alpha, beta, gamma), which are bound  
to each other by disulfide bonds into a cross-shaped molecule  
Comprising one long and three short arms with globules at each  
end.  
CC THE GAMMA-3 CHAIN IS A SUBUNIT OF LAMININ-12.  
CC -!- SUBCELLULAR LOCATION: Extracellular.  
CC -!- TISSUE SPECIFICITY: Strongly expressed in capillaries and  
arterioles of kidney as well as in interstitial Leydig cells of  
testis.  
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT  
WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.  
CC -!- DOMAIN: DOMAIN IV IS GLOBULAR.  
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
CC -!- SIMILARITY: CONTAINS 11 LAMININ EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF083372; AAF08983.1; -.  
DR EMBL; AF079520; AAD29851.1; -.  
DR HSSP; P02468; 1KLO.  
DR MGD; MGI:1344394; Lamc3.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001886; LamNT.  
DR InterPro; IPR000034; Laminin\_B.  
DR InterPro; IPR002049; Laminin\_EGF.  
DR Pfam; PF00052; laminin\_B; 1.  
DR Pfam; PF00053; laminin\_EGF; 10.  
DR Pfam; PF00055; laminin\_Nterm; 1.  
DR PRINTS; PR00011; EGFLAMININ.  
DR ProDom; PD002082; LamNT; 1.  
DR SMART; SM00180; EGF\_Lam; 10.  
DR SMART; SM00281; LamB; 1.  
DR SMART; SM00136; LamNT; 1.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_8.  
DR PROSITE; PS01186; EGF\_2; 2.  
DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 11.  
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.  
FT SIGNAL 1 28  
FT CHAIN 1 28  
FT DOMAIN 29 1581 LAMININ GAMMA-3 CHAIN.  
FT DOMAIN 29 279 LAMININ N-TERMINAL (DOMAIN VI).  
FT DOMAIN 280 335 LAMININ EGF-LIKE 1.  
FT DOMAIN 336 391 LAMININ EGF-LIKE 2.  
FT DOMAIN 392 438 LAMININ EGF-LIKE 3.  
FT DOMAIN 439 488 LAMININ EGF-LIKE 4.

FT	DOMAIN	489	498	LAMININ EGF-LIKE 5 (N-TERMINAL).
FT	DOMAIN	499	584	LAMININ DOMAIN IV.
FT	DOMAIN	685	718	LAMININ EGF-LIKE 5 (C-TERMINAL).
FT	DOMAIN	719	766	LAMININ EGF-LIKE 6.
FT	DOMAIN	767	821	LAMININ EGF-LIKE 7.
FT	DOMAIN	822	877	LAMININ EGF-LIKE 8.
FT	DOMAIN	878	927	LAMININ EGF-LIKE 9.
FT	DOMAIN	928	975	LAMININ EGF-LIKE 10.
FT	DOMAIN	976	1024	LAMININ EGF-LIKE 11.
FT	DOMAIN	1025	1581	DOMAIN II AND I.
FT	DOMAIN	1029	1046	COILED COIL (POTENTIAL).
FT	DOMAIN	1112	1153	COILED COIL (POTENTIAL).
FT	DOMAIN	1208	1231	COILED COIL (POTENTIAL).
FT	DOMAIN	1438	1468	COILED COIL (POTENTIAL).
FT	DOMAIN	1510	1575	COILED COIL (POTENTIAL).
FT	CARBOHYD	128	128	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	304	304	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	337	337	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	640	640	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	849	849	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	991	991	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1162	1162	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1196	1196	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1320	1320	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1514	1514	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	9	9	L -> F (IN REF. 2).
FT	CONFLICT	190	190	P -> T (IN REF. 2).
FT	CONFLICT	195	195	R -> K (IN REF. 2).
FT	CONFLICT	221	221	G -> S (IN REF. 2).
FT	CONFLICT	394	394	C -> R (IN REF. 2).
FT	CONFLICT	471	471	C -> Y (IN REF. 2).
FT	CONFLICT	1150	1150	L -> LDEPOLFSLLK (IN REF. 2).
FT	CONFLICT	1387	1387	H -> Q (IN REF. 2).
FT	CONFLICT	1438	1439	AS -> TI (IN REF. 2).
FT	CONFLICT	1479	1479	V -> I (IN REF. 2).
SQ	SEQUENCE	1581	AA; 172316	MW; 51DFADIF95E6AE81 CRC64;

Query Match 3.1%; Score 173.5; DB 1; Length 1581;  
Best Local Similarity 17.7%; Pred. No. 0.00011;  
Matches 181; Conservative 91; Mismatches 327; Indels 423; Gaps 51;

QY	131	GSGIKFDEWDELPAFESNIATFMDTVVGPSD-----	161
DB	78	GAGLQCQRCDADPCRRHDASYLTDFHSPDDSTWQSPSMAGVQYPTSVNLTLSLGKAY	137
QY	162	-----SRPD-----	185
DB	138	EITYVRLKFTSRPESPAIKRTYASGPWEPIYQYSASCQKTYGRPEGHYLRGDEDERVA	197
QY	186	-CTVSLIYAVHLKKSQYVFEEYQYVDNIIFFETQDQCQEMDTTDDKWKVLTNGEWG	244
DB	198	FTSEFSDISPLNGGNVAFSTLEGRPSAYNFE---ESPVLQEWVTSTDLISLDRLNTFG	254
QY	245	SHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITTEGVAYTSECFFCKPGTFSNKPQS	304
DB	255	DDIFKDPRLVQLQSYYYAVSDFSVGGRC-----KCNHASECEP-----NAAGQ	296
QY	305	ENCQVCPRNTY-----	338
DB	297	LACR-CQHNTTGVDCERCPLFFQDRPWARGTAEDANECLPC---NCSGHSECTFDREL	351
QY	339	-----TERPPC--TTKDYFQ--IHTPCDEEGKTIQIMYKWIPEKICRE--	376
DB	352	YRSTGHGGHCQRCDRHTTGPHCERCEKNYYRWSPKTPC-----QPCDCHPAG	398
QY	377	-----DLTDAIRLPPSGEKKDCPPCPNPGFYNNSSSSCHPC-----PPGTFSDGTKECR	424
DB	399	SLSLQCDNSGVCPCKPTVTGWKCDRCLPGFHSLSGGCRPCACNVAGSLGTCDPRSGNC-	457
QY	425	PCPAGTEPALGFEYK--WNVLPGNMK--TSCFNVNSK-CD-----GM	463
DB	458	PCKENVEGSLCDRCRCPGTENLQPHNPVGCSSCFYGHGSKVCSAAGFQEHHTRSDFRHGA	517







QY 574 QDNRFFNDVMKIY-----SITATNAVD-GVASSCRACALG-----SEQ 611  
Db 1526 CVDTRSGNCYLNIRPRGNDGDTACSNEIGVGVSASCCSLGKAWGTPCELPSVNTSEY 1585  
QY 612 -----SGSSCV-----PCPPGHVIEKETNQ 631  
Db 1586 KILCPGGEGFRNPITVILEIDECQELPGLCQCGKCINTFGSFQCRCPGTGYLNEPTRV 1645  
QY 632 CKE-----CPPDYLSIHQVYGEACIPCGP-----GSKNNQD--HSVCYSDCFF 674  
Db 1646 CDDVNECETPGICGPGT---CYNTVGNVYTCI-CPDYMQVNGNCCMDMRRSICVR----- 1697  
QY 675 YHEKENQILHYDFSNLSSVGLMNGPSFTSKGTYFHFNFNISLCGHEGKKMALCTNNITD 734  
Db 1698 -----NYADNQTCDGELL-----FNMT-----KMCCCSYNIRR 1727  
QY 735 FTVKEI---VAGSDDYTNLVGAFVCQSTIIPSESKGFRAALSSQSIIILADTFIGVTVET 790  
Db 1728 AWNKPCEQCPIPTDEFATLCG-----SORPGF-----VIDIYTGLPVD- 1766  
QY 791 TLKNINIKEDMFPVPSQIP-----DVHFFYKSSSTAT-----TSCIN 827  
Db 1767 -----IDE-----CREIPGVCENGVCINMVGSFRCECPVGFYNDKLLVCEDIDECQN 1814  
QY 828 GR-----STAVKMRNPTKSGAGVIS-----VPSKCPAGTCDGC--TFY 864  
Db 1815 GPVCLRNAECINTAGSYRCD-CKPGYRLTSTGQCNDNRNECQEIPNICSHGQCIDTVGSFY 1873  
QY 865 FL---WESAEACPLCTEHDFHEIE-GACKRG 891  
Db 1874 CLCHTGFKTNEQTMCL--DINECERDAGCNG 1903

RESULT 14

LMB1\_MOUSE STANDARD; PRT; 1786 AA.  
AC P02469;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Laminin beta-1 chain precursor (Laminin B1 chain).  
GN LAMB1-1 OR LAMB-1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87147212; PubMed=3493487;  
RA Sasaki M., Kato S., Kohno K., Martin G.R., Yamada Y.;  
RT "Sequence of the cDNA encoding the laminin B1 chain reveals a  
RT multidomain protein containing cysteine-rich repeats.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:935-939(1987).  
RN [2]  
RP SEQUENCE OF 1292-1786 FROM N.A.  
RX MEDLINE=85051302; PubMed=6209134;  
RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;  
RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of  
RT coiled-coil alpha-helix.";  
RL EMBO J. 3:2355-2362(1984).  
RN [3]  
RP SEQUENCE OF 165-172; 539-547 AND 712-719.  
RC STRAIN=BALB/c; TISSUE=Endothelial cells;  
RX MEDLINE=97363207; PubMed=9219532;  
RA Frieser M., Noeckel H., Pausch F., Roeder C., Hahn A., Deutzmann R.,  
RA Sorokin L.M.;  
RT "Cloning of the mouse laminin alpha 4 cDNA. Expression in a subset of  
RT endothelium.";  
RL Eur. J. Biochem. 246:727-735(1997).  
CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin  
CC is thought to mediate the attachment, migration, and organization  
CC of cells into tissues during embryonic development by interacting  
CC with other extracellular matrix components.

CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three  
CC different polypeptide chains (alpha, beta, gamma), which are bound  
CC to each other by disulfide bonds into a cross-shaped molecule  
CC comprising one long and three short arms with globules at each  
CC end.  
CC THE BETA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-  
CC 2 (MEROSIN), AND LAMININ-6 (K-LAMININ).  
CC -!- SUBCELLULAR LOCATION: Extracellular.  
CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR  
CC COMPONENT).  
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
CC -!- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M15525; AAA39407.1; ALT\_INIT.  
CC EMBL; X05212; CAA28839.1; -.  
CC PIR; A26413; MMSB1.  
CC HSSP; P02468; 1KLO.  
CC MGD; MGI:96743; Lamb1-1.  
CC InterPro; IPR000561; EGF-like.  
CC InterPro; IPR001886; LamNT.  
CC InterPro; IPR002049; Laminin\_EGF.  
CC Pfam; PF00053; laminin\_EGF; 13.  
CC Pfam; PF00055; laminin\_Nterm; 1.  
CC PRINTS; PR00011; EGFLAMININ.  
CC ProDom; PD002082; LamNT; 1.  
CC SMART; SM00180; EGF\_Lam; 11.  
CC SMART; SM00136; LamNT; 1.  
CC PROSITE; PS00022; EGF\_1; 9.  
CC PROSITE; PS01186; EGF\_2; 2.  
CC PROSITE; PS01248; LAMININ\_TYPE\_EGF; 11.  
CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.  
CC SIGNAL 1 21  
FT CHAIN 22 1786 LAMININ BETA-1 CHAIN.  
FT DOMAIN 22 270 LAMININ N-TERMINAL (DOMAIN VI).  
FT DOMAIN 271 334 LAMININ EGF-LIKE 1.  
FT DOMAIN 335 397 LAMININ EGF-LIKE 2.  
FT DOMAIN 398 457 LAMININ EGF-LIKE 3.  
FT DOMAIN 458 509 LAMININ EGF-LIKE 4.  
FT DOMAIN 510 540 LAMININ EGF-LIKE 5 (INCOMPLETE).  
FT DOMAIN 541 772 LAMININ DOMAIN IV.  
FT DOMAIN 773 820 LAMININ EGF-LIKE 6.  
FT DOMAIN 821 866 LAMININ EGF-LIKE 7.  
FT DOMAIN 867 916 LAMININ EGF-LIKE 8.  
FT DOMAIN 917 975 LAMININ EGF-LIKE 9.  
FT DOMAIN 976 1027 LAMININ EGF-LIKE 10.  
FT DOMAIN 1028 1083 LAMININ EGF-LIKE 11.  
FT DOMAIN 1084 1131 LAMININ EGF-LIKE 12.  
FT DOMAIN 1132 1178 LAMININ EGF-LIKE 13.  
FT DOMAIN 1179 1397 DOMAIN II.  
FT DOMAIN 1398 1430 DOMAIN ALPHA.  
FT DOMAIN 1431 1786 DOMAIN I.  
FT DOMAIN 1216 1315 COILED COIL (POTENTIAL).  
FT DOMAIN 1368 1388 COILED COIL (POTENTIAL).  
FT DOMAIN 1448 1778 COILED COIL (POTENTIAL).  
FT DISULFID 271 280 BY SIMILARITY.  
FT DISULFID 273 298 BY SIMILARITY.  
FT DISULFID 300 309 BY SIMILARITY.  
FT DISULFID 312 332 BY SIMILARITY.  
FT DISULFID 335 344 BY SIMILARITY.  
FT DISULFID 337 362 BY SIMILARITY.  
FT DISULFID 365 374 BY SIMILARITY.  
FT DISULFID 377 395 BY SIMILARITY.  
FT DISULFID 398 411 BY SIMILARITY.







FT	DISULFID	244	253	BY SIMILARITY.
FT	DISULFID	260	271	BY SIMILARITY.
FT	DISULFID	265	280	BY SIMILARITY.
FT	DISULFID	282	291	BY SIMILARITY.
FT	DISULFID	298	311	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	322	331	BY SIMILARITY.
FT	DISULFID	338	349	BY SIMILARITY.
FT	DISULFID	343	358	BY SIMILARITY.
FT	DISULFID	360	369	BY SIMILARITY.
FT	DISULFID	375	386	BY SIMILARITY.
FT	DISULFID	380	397	BY SIMILARITY.
FT	DISULFID	399	408	BY SIMILARITY.
FT	DISULFID	415	428	BY SIMILARITY.
FT	DISULFID	422	437	BY SIMILARITY.
FT	DISULFID	439	448	BY SIMILARITY.
FT	DISULFID	455	466	BY SIMILARITY.
FT	DISULFID	460	475	BY SIMILARITY.
FT	DISULFID	477	486	BY SIMILARITY.
FT	DISULFID	493	503	BY SIMILARITY.
FT	DISULFID	498	512	BY SIMILARITY.
FT	DISULFID	514	523	BY SIMILARITY.
FT	DISULFID	530	541	BY SIMILARITY.
FT	DISULFID	535	550	BY SIMILARITY.
FT	DISULFID	552	561	BY SIMILARITY.
FT	DISULFID	568	578	BY SIMILARITY.
FT	DISULFID	573	587	BY SIMILARITY.
FT	DISULFID	589	598	BY SIMILARITY.
FT	DISULFID	605	616	BY SIMILARITY.
FT	DISULFID	627	636	BY SIMILARITY.
FT	DISULFID	643	653	BY SIMILARITY.
FT	DISULFID	648	662	BY SIMILARITY.
FT	DISULFID	664	673	BY SIMILARITY.
FT	DISULFID	680	691	BY SIMILARITY.
FT	DISULFID	685	700	BY SIMILARITY.
FT	DISULFID	702	711	BY SIMILARITY.
FT	DISULFID	718	728	BY SIMILARITY.
FT	DISULFID	723	737	BY SIMILARITY.
FT	DISULFID	739	748	BY SIMILARITY.
FT	DISULFID	755	766	BY SIMILARITY.
FT	DISULFID	760	775	BY SIMILARITY.
FT	DISULFID	777	786	BY SIMILARITY.
FT	DISULFID	793	804	BY SIMILARITY.
FT	DISULFID	798	813	BY SIMILARITY.
FT	DISULFID	815	824	BY SIMILARITY.
FT	DISULFID	831	842	BY SIMILARITY.
FT	DISULFID	836	853	BY SIMILARITY.
FT	DISULFID	855	864	BY SIMILARITY.
FT	DISULFID	871	882	BY SIMILARITY.
FT	DISULFID	876	891	BY SIMILARITY.
FT	DISULFID	893	902	BY SIMILARITY.
FT	DISULFID	909	920	BY SIMILARITY.
FT	DISULFID	914	929	BY SIMILARITY.
FT	DISULFID	931	940	BY SIMILARITY.
FT	DISULFID	947	958	BY SIMILARITY.
FT	DISULFID	952	967	BY SIMILARITY.
FT	DISULFID	969	978	BY SIMILARITY.
Query Match 3.0%; Score 170.5; DB 1; Length 2437;				
Best Local Similarity 19.7%; Pred. No. 0.00034;				
Matches 233; Conservative 102; Mismatches 374; Indels 473; Gaps 71;				
QY	42	QQAAGDL-----PSSSRPLPPCQEKDYHFEYTECDSSGSRWRVAIP-----NS 87		
Db	128	CQPGWSGKTCQLADPCASN-----PCANGG-----QCSAFESHVICTPPNFHGQTCRQD 177		
QY	88	AVDCSGLDPVR-GKEC-----TFSCA-----SGEYLEMKNOVC--SKCGEGTSLGSGI 134		
Db	178	VNECAVSPSPCRNGGTCINEVGSYLCRCPPPEYTGPHCQRLYQCLPSPCRSG----- 229		
QY	135	KFDEWDELPAFESNIATFMDTVVGPSPDSRPDGCNNSSWIP--RGNVIESNRDDCTVSL-- 190		

Db	230	-----GTCVQTSDT-----THTCSCLPGFTGOTCEHNVDDCTQHACE 266
QY	191	-----IYAVHLKSGYVFFEQYVDNNIFFFFEQNDQOEMDT---TTDKWVKL 237
Db	267	NGGPCIDINTYNCHCDK---HWTGOYCTEDV-DECELSPNACONGGTCHNTIGGFHCV 321
QY	238	TDNGEWG-----SHSVMLKSGTNILYWR---TTGIL----- 265
Db	322	CVNGWTGDDCSENIDDCASAAACSHGATCHDRVASFFCECPHGRTGLLCHLDDACISNPCQ 381
QY	266	MGSKA-VKPVLVKNITIEGVAYT-----SECF---PCK-PGTFSNKPGSFNCQVC- 310
Db	382	KGSNCDTNPVSGKAICTPPGYTGSACNQDIDECSLGANPCEHGGRCRLNTKGSFQCKCLQ 441
QY	311	-----PRNTYSEKGAKECIR--CKDDS---QFSG-----SSECTERPP 343
Db	442	GYEGPR--CEMDVNECKSNPCQNDATCLDQIGGFHCICMPGYEGVFCQINSDDCASQPC 498
QY	344	CTTKDYFOI---HTPCDEEGKTQIMYKWIPIKICREDLTDAILRP-----PSGEKKDC 393
Db	499	LNGKCIDKINSFHCECP-----KGFSSLCQVDVDECASTPCKNGAKCTDGPKNYKT 549
QY	394	PPCNPFG-----YNNGSSSCHPCPPGTFSDG-----TKCRP----- 425
Db	550	CECTPGFSGIHCELDINECASS--PCHYGVCRDGVASFTCDCRPGYTGRLCETNINECLS 607
QY	426	-----CPAGTEPALGFYKWNVLPGNMKTSC-FNVGN----- 457
Db	608	QPCRNGGTQDRENAICTCPKGT-----TGVNCEINIDDKRKKPCDY 650
QY	458	SKC-DGMNGWEVAGDHQSGAGGSDN-DYLILN-----LHIPGFKPPT 498
Db	651	GKCIDKINGECVCEPGYSGSMCNINIDDCALNPCHNGGTCIDGVNSFTCLCPDGFDRDAT 710
QY	499	SMTGAT-----GSELGRITTFVFETLCSA-----DCVLYFMVDINRKSTNVVSWGG 544
Db	711	CLSQHNECSSNPCIHGSCLDQIN-SYRCVCEAGWMGRNC---DININECLSNPCVN-GG 764
QY	545	TKEKQAYTHIIFKNATFTTWAQFQRTNOGQDNRRTINDMKVIYSITATNAVGVASSCRA 604
Db	765	TCKDMTSGYLCTCRAGFS-----GPNCOMNINECASNPCLNQGSCIDDDVAGFKCN 814
QY	605	CALGSEQSGSSC---VPCPPGHYIEKETNOCKE-----CPPDTYLSIHQVYGKE 650
Db	815	CML--PYTGEVCENVLAPCSRPC--KNGGVCRESEDFQSFSCNCPAGWQGTCEVDINE 870
QY	651	ACI--PCGPGSKNQDHSVYSDCFYHEKENOILHYDFSNLSSVGLMNGPSTFSKGTK 708
Db	871	-CVRNPCTNG-----GVC-----ENLRGGFQCRCPNGFTG---- 899
QY	709	YPHFENISLCGHEGKKMALGTNNITDFTVKEIVAGS--DDYTNLVGAFCVQOSTIIPSESK 766
Db	900	-----ALCENDIDDCENPCSNNGGVQCQDRVN---GFVCVCL-----A 933
QY	767	GFRALSSQSIILADTFIGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSSTATTSCT 826
Db	934	GERGERCAEDI-----DECVSAPC-----RNGGNC'TDCV 962
QY	827	NGRSTAVKMRNPTKSGAGV---ISVP---SKC-PAGTC-DGCTFFYFLWESAEACPLCT 877
Db	963	NSYT-----CSCPAGFSGINCENPTDCTESSCFNGGTCVDGSISSF---SCVCLPGFT 1012
QY	878	-----EHDFFEIE-----GACKRGFQ-----ETLYVWNEPKWCIKIGISL 911
Db	1013	GNVCOHDVNECDSRPCQNGGSCQDGYGYTKYKTCPTHGYTGLNCQSLVRWCDSSPCCKNGGSC 1072
QY	912	PEKLA-TCETVDVFWL-----KVGAGVGAFATVLLVLTLC 945
Db	1073	WQQGASFTCCQASGWTGIYCDVPSVSCEVAARQQGVSAVLIC 1114







GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: May 12, 2003, 12:50:23 ; Search time 43.8668 Seconds  
(without alignments)  
4823.927 Million cell updates/sec

Title: US-10-073-333A-4  
Perfect score: 5681  
Sequence: 1 MLFRARGPVRGRGWRPAEA.....KEKEDHFESVQLKTSRSPNI 1027

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

SUMMARIES				% Query Match		Length DB		ID		Description	
Result No.	Score	Match	Length	DB	ID						
1	2624	46.2	493	4	Q96DP2					Q96dp2 homo sapien	
2	1753	30.9	580	4	Q9P2M2					Q9p2m2 homo sapien	
3	1305	23.0	1019	5	Q9NA40					Q9na40 caenorhabdi	
4	888.5	15.6	300	11	Q8R215					Q8r215 mus musculu	
5	225.5	4.0	1074	5	Q964D1					Q964d1 entamoeba h	
6	213.5	3.8	1101	5	Q964D2					Q964d2 entamoeba h	
7	207.5	3.7	704	5	Q9U048					Q9u048 giardia lam	
8	203	3.6	709	5	Q9XTJ7					Q9xtj7 giardia lam	
9	201	3.5	709	5	Q97444					Q97444 giardia lam	
10	199.5	3.5	667	5	Q95WU1					Q95wu1 giardia lam	
11	199.5	3.5	2759	5	O45614					O45614 caenorhabdi	
12	199.5	3.5	3102	5	Q9T2R4					Q9t2r4 caenorhabdi	
13	197.5	3.5	3087	5	Q8WRD2					Q8wrd2 plasmodium	
14	193.5	3.4	5374	11	Q99ND0					Q99nd0 mus musculu	
15	192.5	3.4	719	5	Q9U021					Q9u021 giardia lam	
16	190	3.3	769	5	Q24971					Q24971 giardia lam	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

17	188.5	3.3	999	4	Q9NQ36	Q9nq36	homo sapien
18	187.5	3.3	1557	5	Q8WPK9	Q8wpk9	oikopleura
19	187	3.3	1723	5	Q8WRD0	Q8wrd0	plasmodium
20	186	3.3	1007	13	Q90ZN3	Q90zn3	gallus galli
21	185.5	3.3	719	5	Q9U019	Q9u019	giardia lam
22	184.5	3.2	2189	5	Q9BI05	Q9bi05	eimeria ten
23	184.5	3.2	2534	5	Q8WRD1	Q8wrd1	plasmodium
24	183	3.2	1372	5	P9I526	P9i526	caenorhabdi
25	182	3.2	2931	5	Q9W2C6	Q9w2c6	drosophila
26	181.5	3.2	1551	5	Q9NGV4	Q9ngv4	drosophila
27	181.5	3.2	3396	5	Q9VM55	Q9vm55	drosophila
28	180.5	3.2	1289	5	Q8SSS3	Q8sss3	dictyosteli
29	179.5	3.2	718	5	Q9BI07	Q9bi07	entamoeba h
30	179.5	3.2	3567	11	Q9ES77	Q9es77	mus musculu
31	178.5	3.1	3704	5	P9I904	P9i904	caenorhabdi
32	175	3.1	2906	11	Q9WUH9	Q9wuh9	rattus norv
33	174.5	3.1	548	5	Q9GQ45	Q9gg45	giardia lam
34	174.5	3.1	597	5	Q07317	Q07317	giardia int
35	174.5	3.1	1316	4	Q96JU7	Q96ju7	homo sapien
36	173.5	3.1	468	5	Q9BMH3	Q9bmh3	ichthyophth
37	173	3.0	667	5	Q9XTK3	Q9xtk3	giardia lam
38	173	3.0	1045	5	Q8T3A6	Q8t3a6	caenorhabdi
39	173	3.0	1070	5	Q8T3A7	Q8t3a7	caenorhabdi
40	173	3.0	1111	5	Q9XWD6	Q9xwd6	caenorhabdi
41	172.5	3.0	3857	11	O88840	O88840	mus musculu
42	171.5	3.0	2809	4	Q96JP8	Q96jp8	homo sapien
43	169.5	3.0	739	5	Q9GS24	Q9gs24	giardia lam
44	168	3.0	594	5	Q24970	Q24970	giardia lam
45	168	3.0	1664	5	Q9TVQ2	Q9tvq2	caenorhabdi



Db	64	SFNCQVCPRNTYSEGAKECIRCKDDSQFSEGSSECTERPPCTTKDYFQIHTPCDEEGK	123
QY	362	TQIMYKWIPEKICREDLTDAIRLPPSGEKKDCPCNPGFYNGSSSCHPCPPGTFS	421
Db	124	TQIMYKWIPEKICREDLTDAIRLPPSGEKKDCPCNPGFYNGSSSCHPCPPGTFS	183
QY	422	ECRPPCAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSD	481
Db	184	ECRPPCAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSD	243
QY	482	NDYLILNLHIPGFKPPTSMTGATGSELGRITFVFETLCSADCVLVYFMVDINRKSTNVVES	541
Db	244	NDYLILNLHIPGFKPPTSMTGATGSELGRITFVFETLCSADCVLVYFMVDINRKSTNVVES	303
QY	542	WGGTKEKQAYTHIIFKNATFTFTWAFQRTNQGDNRREFINDMVKIYSITATNAVDGVASS	601
Db	304	WGGTKEKQAYTHIIFKNATFTFTWAFQRTNQGDNRREFINDMVKIYSITATNAVDGVASS	363
QY	602	CRACALGSEQSGSSCVCPPPGHYIEKETNQCKECPDPTYLSIHQVYGKEACIPCGPGSKN	661
Db	364	CRACALGSEQSGSSCVCPPPGHYIEKETNQCKECPDPTYLSIHQVYGKEACIPCGPGSKN	423
QY	662	NQDHSVYCSDCFFYHEKENQILHYDFSNNLSSVGLMNGPSTSGTKYFHFFNISLGGHE	721
Db	424	NQDHSVYCSDCFFYHEKENQSLHYDFSNNLSSVGLMNGPSTSGTKYFHFFNISYVGMR	483
QY	722	GKK 724	
Db	484	GRR 486	
RESULT 2			
QY	Q9P2M2	PRELIMINARY;	PRT; 580 AA.
AC	Q9P2M2		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	KIAA1324 protein (Fragment).		
GN	KIAA1324.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=BRAIN;		
RX	MEDLINE=20181126; PubMed=10718198;		
RA	Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;		
RT	"Prediction of the coding sequences of unidentified human genes.XVI.		
RT	The complete sequences of 150 new cDNA clones from brain which code		
RT	for large proteins in vitro."		
RL	DNA Res. 7:65-73(2000).		
DR	EMBL; AB037745; BAA92562.1; -.		
FT	NON_TER		
SQ	SEQUENCE 580 AA; 63365 MW; 5E63C19265EC8E5C CRC64;		
Query Match			
Best Local Similarity 30.9%; Score 1753; DB 4; Length 580;			
Matches 307; Conservative 103; Mismatches 157; Indels 4; Gaps 2;			
QY	440	WNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTS	499
Db	1	WNTLPTNMETVLSGINFYKMGTMGWEVAGDHIYTAAGASDNDFMILTLLVVPGRPPQS	60
QY	500	-MTGATGSELGRITFVFETLCSADCVLVYFMVDINRKSTNVVESWGGTKEKQAYTHIIFKN	558
Db	61	VMADTENKEVARITFVFETLCSVNCLEYFMVGVNSRTNTPVETWKSGKQSYTHIIEEN	120
QY	559	ATFTFTWAFQRTNQGDNRREFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVP	618
Db	121	TTTSFTWAFQRTTFEASRKYTNDVAKIYSINVTNMVNGVASYCRPCALEASDVGSCTS	180

QY	619	CPPGHYIEKETNQCKECPDPTYLSIHQVYGKEACIPCGGSKNNQDHSVYCSDCFFYHEK	678
Db	181	CPAGYIDRDSGTCHSCPNTILKAHQPYGVACVPCGPGTKNNKIHSCLYNDCTFSRNT	240
QY	679	ENQILHYDFSNSVGLMNGPSTSGTKYFHEFNISLCHGHEGKKMALCTNNITDFTVK	738
Db	241	PTRTFNYNFSALANTVTLAGGPSTSGTKYFHEFTLSLCGNQGRKMSVCTDNVTDLRIP	300
QY	739	EIVAGSDDYTNLVGAFCQSTIIPSESKGFRAALSSQSIILADTFIGVTVETILKNINIK	798
Db	301	E--GESGFSKITAYVQAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSP	357
QY	799	EDMFPVPTSQIPDVHFFYKSSSTATTSINGRSTAVKMRCPNPKSGAGVISVPSKCPAGTC	858
Db	358	AELFHLESGLIPDVIFVRSNDVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPSTCSDGTC	417
QY	859	DGCTFYFLWESAEACPLCTEHDHFHEIEGACKRGFQETLYVWNEPKWCIKGISLPEKKLAT	918
Db	418	DGCNHFELWESAAACPLCSVADYHAIVSSCVAGIQKTYVWREPKLCSGGISLPEQRTI	477
QY	919	CETVDFWLKVGAGVGAFTAVLLVALTCYFWKKNQKLEKYSKLVMTTNSKECELPADSC	978
Db	478	CKTIDFWLKVGISAGTCTAILLTVLTCYFWKKNQKLEKYSKLVMTATLKDCDLPADSC	537
QY	979	ATMEGEDNEEVVYSNKQSLGKLKSLATKE 1009	
Db	538	ATMEGEDVEDDLIFTSKKSLFGKIKSFTSKQ 568	
RESULT 3			
QY	Q9NA40	PRELIMINARY;	PRT; 1019 AA.
AC	Q9NA40		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
DE	Y73F8A.5 protein.		
GN	Y73F8A.5.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Matthews L.;		
RL	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99069613; PubMed=9851916;		
RA	none;		
RT	"Genome sequence of the nematode C.elegans: A platform for		
RT	investigating biology."		
RL	Science 282:2012-2018(1998).		
DR	EMBL; AL132862; CAB70224.1; -.		
DR	InterPro; IPR002049; Laminin_EGF.		
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.		
SQ	SEQUENCE 1019 AA; 112378 MW; 6EB4094722B707E7 CRC64;		
Query Match			
Best Local Similarity 23.0%; Score 1305; DB 5; Length 1019;			
Matches 316; Conservative 175; Mismatches 398; Indels 118; Gaps 40;			
QY	61	CQEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC	120
Db	25	CTDQDYEFYTYTNCDENGERWRVAVPRGGMQCSNLPTRRGLNCSFSCPEGHVLDLDSQHC	84
QY	121	SKCGEGTYSLSGSGTKFDEWDELPAQFSNIATFMDTVVGPDSRDPG-----C-N 168	
Db	85	RPCNPGFFSLGGGIRYEEFVTLPSGFS-----VDNMDSNPDAQFSNRQSQVVECPK 135	
QY	169	NSSWIPRGNYESNRDDCTVSLIYAVHLKKSIFYFEYQYVDNN--IFFEFTIQNDQCQE 226	

Db 136 EAGVVVKDGLIYIPTCVSKLSFSAANLVRPGSVEFTYRMPNRNRLSMQVDIRNEQCQS 195

QY 227 M-DTTTDKWKVLT-----DNGEGSHSVMLKSGTNILYWRRTGILMGSKAVKPVLVK 277

Db 196 YNDVAKSMFLKYTKKEKDEEKNKGDRKRRIELKSGANVISWIIQNMGYQASNOPIHID 255

QY 278 NITIEGVAYTSECFCKPGTFSNKPGSFNCQVCPRNTYSEKAKECIRCDDSQFSG--S 335

Db 256 RIDVLGLAFTROCTACPPGT-SSPGSAECIPCSPGFSFSSKSGQCGR-C-PESQYSGFKS 313

QY 336 SECTERPPCTTKDYFOIHTPCDEGKTQIMYKWIIEPKICREDLTDAILPPSGEKKDCPP 395

Db 314 EKCIDRPPCRVSDYYPVREPC-TNGSSRAVYKKVLPISICRDDLPSATKLPPTPWKTCPK 372

QY 396 CNPGFYNNGSSSCHPCPPGTFSDGTEKCRPCPAGTEPALGFEYKWNVLPGNMKTSCFNV 455

Db 373 CNPGMEKNKLGVECFCKKDYFSDG-NSCSRCPVDTPVNYGLOQYQNWVLPKLSRCEYI 431

QY 456 GN---SKCDGMNGWEVAGDHIQSGAG--GSDNDYLILNLHIPGF-----KPPTSMTGA 503

Db 432 SEDVATACNIGDAWIPSGDSLISAPSLGIAFE-LILSID-EGFWNPLAPKSKTMK-- 487

QY 504 TGSELGRITVFETLCSAD-CVLYFMVDIN-----RKS-TNVVESWGGTKEKQAYTHII 555

Db 488 --VPVAQTVIVETSCADESCALYFIEDMSAGIKGQRESEFYHFLAAFNAGSNKRWSHTV 545

QY 556 FKNATFTFTWAFQRT--NOQQDNRFFINDVMKIYSITATN-----AVDGVASSCRACALG 608

Db 546 TKNTPARFMVAFLRSGVSSGDDK--ISDETRIYAINVTNVGHRGGGGASQCLTCP-- 600

QY 609 SEQSGSCVPCPPGHIYIEKETNOCKECPDPTYLSI-HQVYKGEACIPCPCGPGSKNNQDHSV 667

Db 601 HTAGETCVPCPAGNYMHEVTKLCVSCPLNTIINVTSSRVGVKSCVPCGQGLTSDNGVSC 660

QY 668 CYSDCFFYHE-----KENQILHYDFS-----NLSSVGLMNGPSPFTSKGTYFHFENI 715

Db 661 TAMGKIQLNQIGGKGNDSSTYDFSPFVGRSWNISGV-----RVFSREGAAYYHFFSV 714

QY 716 SLCGHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFCVQSTIIPSESKGFRAALSSQ 775

Db 715 SLFPPNIK---CQEQFDNFMIGIL--DQDKENVEG-LACRVATALPTSSNRSKTAAYT 767

QY 776 SIILADTFIGTVETTLKLNINIKEDMFPVP---TSQIPDVHFFYKS-STATTSCINGRS 830

Db 768 PLLLAGRLDSISLNRKHGNSLSDEVLEYDSDHNTSHPLDVFFWFEPVSTISPACPNQ 827

QY 831 TAVKMRNPTKSGAGVISVPSKCPAGTCDCGCTFYFLWESAEACPLCTEHDFFEIEGACKR 890

Db 828 LVVVARCVPTKKQM-EMRLPHNCPDGTCDGCLFVIIMETAQACPVCESNDYETINGECLN 886

QY 891 GFQETLYVWNEPKWCIKIGISLPEKKLATCETVDFWLKVGAGVGAFVALL-VALTCYFPWK 949

Db 887 G-KOTIHSIPKKHCVITGAASQSKEVA-CSAFTAFQRAILSILVLSMIFLSIGFVC-ICR 943

QY 950 KNOKLEYKYSKLVMTTNSKECELPAAADSCAIMEGEDNEE---EVVYS 993

Db 944 RNRLEYKYTRLI---ESHTGELPAVETCGLDEDEDDDELQDRVIFS 987

RESULT 4

Q8R215

ID Q8R215 PRELIMINARY; PRT; 300 AA.

AC Q8R215;

DT 01-JUN-2002 (TremBLrel. 21, Created)

DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)

DE Hypothetical 32.6 kDa protein (Fragment).

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=COLON;

RA Strausberg R.;

RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC022655; AAH22655.1; -

KW Hypothetical protein.

FT NON\_TER 1

SQ SEQUENCE 300 AA; 32642 MW; 6455109C054B6CB4 CRC64;

Query Match 15.6%; Score 888.5; DB 11; Length 300;

Best Local Similarity 53.6%; Pred. No. 2.3e-67;

Matches 158; Conservative 52; Mismatches 82; Indels 3; Gaps 1;

QY 728 CTNNITDFTVKEIVAGSDDYTNLVGAFCVQSTIIPSESKGFRAALSSQSIILADTFIGVT 787

Db 1 CTDNVTDLRIPDGEAG--FSKSVTAYVCQVWIIPSEVMGYKAGVSSQPVSLADRLVGVS 57

QY 788 VETTLKLNINIKEDMFPVPTSQIPDVHFFYKSSSTATTSCINGRSTAVKMRNPTKSGAGVI 847

Db 58 TDMTLEGIVSPVELFHPETSGIPDIVFFRSNDVTQSCSSGRSTIRLCNPMKAAPGTL 117

QY 848 SVPSKCPAGTCDCGCTFYFLWESAEACPLCTEHDFFEIEGACKRGFQETLYVWNEPKWCIK 907

Db 118 RLPSMCSGTCDCGCFNFHFLWESVAACPLCSASDYHTFVSSCVAGIQKTTYMWREPCKLCSG 177

QY 908 GISLPEKKLATCETVDFWLKVGAGVGAFVALLVALTCYFWKKNOKLEYKYSKLVMTTNS 967

Db 178 GISLPEQRTVICTIDFWLKVGISAGTCTAILLTVLTCYFWKKNOKLEYKYSKLVMNATL 237

QY 968 KECELPAAADSCAIMEGEDNEEVVYSNKSQSLGKLSLATKEKEDHFEVQLKTS 1022

Db 238 KCDLPAADSCAIMEGEDVEDDFITSKSLFGKIKSFTSKRTPDGFDSVPLKTS 292

RESULT 5

Q964D1

ID Q964D1 PRELIMINARY; PRT; 1074 AA.

AC Q964D1;

DT 01-DEC-2001 (TremBLrel. 19, Created)

DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)

DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)

DE Gal/GalNAC lectin Igl2 (Fragment).

GN IGL2.

OS Entamoeba histolytica.

OC Eukaryota; Entamoebidae; Entamoeba.

OX NCBI\_TaxID=5759;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=HMI:IMSS;

RX MEDLINE=21391855; PubMed=11500468;

RA Cheng X.J., Hughes M.A., Huston C.D., Loftus B., Gilchrist C.A.,

RA Lockhart L.A., Ghosh S., Miller-Sims V., Mann B.J., Petri W.A. Jr.,

RA Tachibana H.;

RT "Intermediate Subunit of the Gal/GalNAC Lectin of Entamoeba

RT histolytica Is a Member of a Gene Family Containing Multiple CXXC

RT Sequence Motifs.";

RL Infect. Immun. 69:5892-5898(2001).

DR EMBL; AF337951; AAK92362.1; -

DR InterPro; IPR000561; EGF-like.

DR PROSITE; PS01186; EGF\_2; UNKNOWN\_1.

KW Lectin.

FT NON\_TER 1

FT NON\_TER 1074

SQ SEQUENCE 1074 AA; 116782 MW; 1552E2D714EB450F CRC64;

Query Match 4.0%; Score 225.5; DB 5; Length 1074;

Best Local Similarity 19.5%; Pred. No. 5.1e-10;

Matches 203; Conservative 117; Mismatches 392; Indels 329; Gaps 54;

QY 107 CASGEYLEMKNQVCSKCGEGTYSLGSGIKFDEWDELPAFNSNIATFMDTVVGPDSRDPG 166

Db 13 CAS-----VSNGACASCDEG-YELKT-----ESGSGSTQKCTLKEETCKSAFSYVDG 58

QY 167 CNSSWIIPRGNIESNRDDCTVSLIYAVHLKKGYSVFVEYQVVDNNIFFEFI--QNDQC 224





```
Db 571 SKVVCSEYENIQGVTRNKENEACINDGY-----KE-----GPNA 607
QY 703 TSKGTKYFHFNFISLCGHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFV-----CQ 757
Db 608 EDKKKSCAQLNN--NCKKEGK-----YEISDGFV--TCLDCDDSAIYVGSQVGACTQCS 657
QY 758 STIIPSESKGFRAALSSQSIIILADTFIGVTVETTLKNNIKEDMFPVPTSQIPDV---HF 814
Db 658 PNAFKDENNKQLCSTKQSQYGHCAACSAFACITCEDINL-----ILTGEKPCTVCKDG 711
QY 815 FYKSSTAT-----TSC--INGRSTAVKMRC-NPTKSGAGVISVPS-KCPA-----855
Db 712 FYQIENATDGVYCSPCPAKCKTCKYNTTSKKVECVCTEQRLKDIIKAPECACTGTVOLE 771
QY 856 -GTCDCGT-----FYFLWESAEA-----CPLCTEHDFHEI 884
Db 772 NGTCQSCSDLKYPGCKKTDSCNVDSRTGTFIYATECSDFSGRSPYSNCTTCTKSNIYPK 831
QY 885 EGACKRGFQETLYVMNEPKWCIIKISLPEKKLATCETVDFW-----LKVGAAGVGAFTAV 938
Db 832 EG-----EKNGCAK-----CDDKCATCSDKDTCLTCADPLKVGSKCDG----869
QY 939 LLVALTCYFWKKNOKLEYKYSKLVMTTNSKECELPAAADSCAIMEGE 984
Db 870 ---CKTGYMNSG---ECK---PCTNHCSECS--SAAECTVCESD 903

RESULT 7
Q9U048
ID Q9U048 PRELIMINARY; PRT; 704 AA.
AC Q9U048;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Variant-specific surface protein.
GN VSP417-6.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AD-1;
RA Ey P.L., Darby J.M.;
RT "A new member of the vsp417 subfamily of variant-specific surface
RT protein (VSP) genes in Giardia intestinalis.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF065606; AAF02907.1;
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR005127; Giardia_VSP.
DR Pfam; PF03302; VSP; 2.
DR SMART; SM00261; FU; 4.
SQ SEQUENCE 704 AA; 71677 MW; 7E5AE1245AD4FD45 CRC64;

Query Match 3.7%; Score 207.5; DB 5; Length 704;
Best Local Similarity 20.8%; Pred. No. 9.5e-09;
Matches 199; Conservative 88; Mismatches 353; Indels 315; Gaps 54;

QY 38 ALAGCQAAWAGDLPSSSRPLPPCQEKDYHFEY-TECDSSGSRWRVAIPNSAVDCSGLPD 96
Db 15 ARAACQADGSGSDGCKTCGVTIGQE-----QYCSECNGA-----NYA-----52
QY 97 PVRGKECTFSCASGEYLEM---KNQ-VCSKCG-----EGTYSLSGGIKFDEWDELPAG 145
Db 53 PVNG-QCADVNAEQSQKTLCPAKGQKCTQCGGASFMYKDGCS-----95
QY 146 FSNIAFMDTVVGPS---DSRPDGCNNSSWIPRGNYESNRDDCTVSLIYAVHLKSGYV 202
Db 96 -----KDTAPGQSMCTQASEGKCTEAPGYFLNPLRANTKDSVYVSCSDTTGTFTDSGKT 148
QY 203 FFEYQYVDNIIFFEFFIQNDQCQEMDTTDDKWVKLTDNGEWGSHSVMLKSGTNILYWRRT 262
Db 149 YRGVQH-----CERCDCGAA-----LTDA--GGDAKCTRCGQD-KYLATT 185
```

```
QY 263 GILMGSKAVKPVLVKNITIEGVAYTSECFFCKPGTFSNKPGSFN---CQVCPRNTYSEKG 319
Db 186 GTCG-----EG-----CTPDTEFSKEDSDNGKRCFACGDVT---TG 218
QY 320 AKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIIEPKICREDLT 379
Db 219 VASCEKCTPPSPDQAKPACTK---CGGNYY--LKTAAD--GTTTC---VEQSACSPD-S 266
QY 380 DAIRLPPSGEKKDCPPCNPFGFYNNGSSSCHPCPCPTFSDGTKECRPCPAGTEPALGFEYK 439
Db 267 FPVENSQSGNR--CVLCGDA-ANGGVDKCAACTP-----TDQGRIAPTITCIACNGYK 317
QY 440 WNVLPGNMKTSCFNVCNSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGPKPPTS 499
Db 318 ----PSADKTTCEAVSNCKTPGCKACSNEGKENEVECTDCDSSYL-----TPTS 362
QY 500 MTGATGSELGRITFVFETLCSADCVLYFMVDINRKNSTNVVESWGGTKEKQAYTHIFKNA 559
Db 363 Q-----CIDSCA-----KIGNYYGATEGAKK-----ICKEC 388
QY 560 TETFTWAFORTNOGQDNRFRINDMVKIYSITATNAVGVASSCRACALGSESGSSCVPC 619
Db 389 TA-----ANCKTCDGQ-----GQCACSDGFYKNGDACSPC 419
QY 620 PPGHYIEK-----ETNQCKECPDLYLSIHQVYGKEACI-PCGPGSKNNQDHSHVCYSDCF 673
Db 420 ---HESCKTCSAGTASDCTECPTGKALR---YGGDTKGTCGEGCTTGQGSAC-----467
QY 674 FYHEKENQILHYDFSNLSSVGLMNGSPFTSKGTKYFHFENISLCGHEGKKMALCTNNIT 733
Db 468 -----KTCGLTIDGASYCSECATTEYPQNGVC--APKASRATPTCN 507
QY 734 DFTVKEIVAG--SDDYTNLVGAFCVQCSTIIPSESKGFRAALSSQSIIILADTFIGTVVETT 791
Db 508 DSPIONGVCGTCADNYFKMNGG--CYETV--KYPGKTVICISAPN-----GGTCQKA 554
QY 792 LKNINIKEDMFPVPTSQIPDVHFFYKSSSTATTSCTSCING--RSTAVKMRCN---PTKSGAGV 846
Db 555 ADGYKLDSGTLTVCSGCKEC-----TSSTDCTTCLDGYVKASASACTKCDSSCETCNGA--608
QY 847 ISVPSKCPAGTCDGC-TFYFLWESAE-ACPLCTEHDFHEIEGACKRGFQETLYVMNEPKW 904
Db 609 -----ATTCKACATGYKYKTASGEGACTSC-ESDSNGVTGI-----KG 644
QY 905 CIKGISLPEKKLAT--CETVDFWLKVGAGVGAFATVL-----LVALTCYFW 948
Db 645 CL-NCAPPSSTGSLVLCYLQNTNKSGLSAGAIAGISVAVIVVWAGLVGLCWWF 698

RESULT 8
Q9XTJ7
ID Q9XTJ7 PRELIMINARY; PRT; 709 AA.
AC Q9XTJ7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Variant-specific surface protein (Fragment).
GN VSP417-6.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRIS-136;
RX MEDLINE=99026095; PubMed=9806870;
RA Ey P.L., Darby J.M.;
RT "Giardia intestinalis: conservation of the variant-specific surface
RT protein VSP417-1 (TSA417) and identification of a divergent homologue
RT encoded at a duplicated locus in genetic group II isolates.";
RL Exp. Parasitol. 90:250-261(1998).
DR EMBL; U89266; AAD03483.1;
DR InterPro; IPR002174; Furin-like.
```









Db 873 -----CSDGFEDPLTGKIECTCNGNIDPMIGNCDS-----ETGKCLKCI 914  
QY 124 GEGTYSLGSIGIKFDEWDELPAAGFSNIATFMDTVVGPDSRDPDGCNNSSWIPRGNYIESNR 183  
Db 915 GHTTGDSCESCSEKHHW-----GNAQLHT-----CKPCGCHTQGA VNPQCSENGE 959  
QY 184 DDCTVSLIYAVHLKKSQYVFFQYVYVDDNNIFFEFFIQDQCQEMDITTDKWKVLTNDGEW 243  
Db 960 CECKENYIGA-----QCDCKENHGDVENGCPACDCNDT 993  
QY 244 GSHSVMLKSGTNILYWRRTGILMGSKAVKPVLVKNITIEGVAYTSECFCCKPGTFSN--- 300  
Db 994 GS-----IGSDCDQ-----VSGQC-NCKQGVFGKQCD 1019  
QY 301 --KPGSFN-----CQVCPRNTYSEKGAKECIRCKDDSQFSGSSECTERPPCT-----TKD 348  
Db 1020 QCRPSYFNFTDAGCQFCHCNIY---GSIEDGKC---DQTGKCECRENVGTCMEKCADG 1073  
QY 349 YFQIHT-----PCDEEGKTQIMYKWIIEPK-ICREDLTDAILRPPSGEKKDCPPCNPGE 400  
Db 1074 YFNITSGDGCEDCGCDPTGSEDVSCNLVTGQCVCCKPGVT-----GLK--CDSCCLPNE 1123  
QY 401 Y---NNGSSSCHPCP-PGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVG 456  
Db 1124 YGLTSEGCTECPAPGVQVCDPIDGSCVCPNT-----VEMCENCCTT-- 1167  
QY 457 NSKCDGMNGWEVAGDHIQSGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITFVFE 516  
Db 1168 -----NAW-----DYHPLN-----GCK----- 1179  
QY 517 TLCSADCVLYFMVDINRKSTNVVESWGCTKEKQAY-----THIIFKNATFTTFAFQ 568  
Db 1180 -LC--DC-----SDIGSDGMCNTFTGQCKCKAAVYVGLKCDLCTHGFNFPT----- 1223  
QY 569 RTNQGDNRFRINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGHYIEKE 628  
Db 1224 -----CNEICECP-----CKKNVHGTK-CDQCGEGTFLDSSNLKGCTECF--CFNRTSNCEQSD 1297  
QY 674 -----FYHEKEN-OILHYDFSNLSS-----VGLSMNGPS 701  
Db 1298 LVWQQYAEADRRRAVFOEPWEFYTKKHNLNLLREKPSHFSNYPTDATPLYWLPSTMLGDR 1357  
QY 702 FTS-----KGTKYF-HFFNISLCGHEGKKMALCTNNITDFTVK 738  
Db 1358 TASYNGFLRFKIWNEDRRGLHGRPDQOQYFRHFPQVFIIFGNRIELEHIPMEINDDGIY 1417  
QY 739 EI-----VAGSDDYT-----NLVGAFCVQSTIIPSESKGFRAALSSQSIL 779  
Db 1418 KIRLHESWRVRHSPELTLTRKQMMVALQDTQGIYIRGTYPARGDAINIQEVSLDVAV 1477  
QY 780 ADTFIGVTETTLKNINIKEDM-----FPVPTSQIPDVHFFYKSS----- 819  
Db 1478 PESKIVAGLSTT-KAIGVEKCLGCPQGYTGLSCQNPVEGYRKKHREYLNQADDIALIGW 1536  
QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPKCPAGT-----CDGCTFYFLWESAE---- 871  
Db 1537 SEPCSC-HGHS---QTCNPD-----TSVCTDCEHNTFGDFCEHCLPGYIGDAREGGAN 1585  
QY 872 -----ACPL-----CTEHDFHEIEGACKRGFOETLYVWNEPKWCIKG-ISLPEKK 915  
Db 1586 ACTKACAPLVENSFSDSCVAVDHRGVYVNCCKPG-----YTGOYCETCVAGYIGDPQHI 1640  
QY 916 LATCETVDFWLKVAGVGAFTAVL-----LVALTCYFWKKNQKLEYKYSKLVMTTNS 967  
Db 1641 GGTCSPCDCH-PDGLSHGACNPLSGQCECKPGVTGRTCSMCQE---RHAFINRVCTSCD 1695  
QY 968 KECELPAADSCAIMEGEDNEEYVSNKQSLGLKLSLATK 1008

Db 1696 QGCYLP-----LMETMDTMEE--HLGRONFSG-LKPIPWK 1727  
RESULT 13  
Q8WRD2  
ID Q8WRD2 PRELIMINARY; PRT; 3087 AA.  
AC Q8WRD2;  
DT 01-MAR-2002 (TremBLrel. 20, Created)  
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE Cysteine repeat modular protein 1 PbCRM1.  
OS Plasmodium berghei (strain Anka).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ANKA;  
RA Thompson J.;  
RT "Plasmodium berghei Cysteine Repeat Modular Protein 1: PbCRM1";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF449196; AAL47156.1; -;  
DR InterPro; IPR000345; CytC\_heme\_bind.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR001664; IF.  
DR InterPro; IPR000001; Kringle.  
DR InterPro; IPR000589; Ribosomal\_S15.  
DR Pfam; PF00051; kringle; 1.  
DR SMART; SM00130; KR; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
DR PROSITE; PS01186; EGF\_2; UNKNOWN\_3.  
DR PROSITE; PS00226; IF; UNKNOWN\_1.  
DR PROSITE; PS00070; KRINGLE\_2; 1.  
DR PROSITE; PS00362; RIBOSOMAL\_S15; UNKNOWN\_1.  
SQ SEQUENCE 3087 AA; 362575 MW; 38BDD25AB65FD43D CRC64;  
Query Match 3.5%; Score 197.5; DB 5; Length 3087;  
Best Local Similarity 20.0%; Pred. No. 5.7e-07;  
Matches 197; Conservative 127; Mismatches 322; Indels 337; Gaps 55;  
QY 102 ECTFSCASGEYLEMKNOVCSCGEGTYSLGSGIKFDEWDELPAAGFSNIATFMDTVVGPSPD 161  
Db 1317 EFSFACEDGHY--HKHNKCHPCCKGYNNLNEIK-----QNN 1351  
QY 162 SRPDGC---NNSSWIPRGNYIESNRDDCTVSLIYAVHLK-----SGYVFFEQY 208  
Db 1352 EKYKKCTACQONRTTLDKKEKEKN--CLCDLGYE-YIKNPNPNRNFICSPCVYGEYKD 1407  
QY 209 VDNNIFFEFFIQDQCQEMDITTDKWKVLTNDGEWGSVLMKSGTNILYWRRTGILMGS 268  
Db 1408 V-----ISNELK-----GIVCKKNASYL-----ILDK 1430  
QY 269 KAVKPVLVKNITIEGVAYTSECFCCKPGTF--SNKPGSFNCQVCPRTYSEKGAKECIRC 326  
Db 1431 KSVDP-----SQCL-CNGGYLYLRDKNNNEICTKINNHY-----C 1465  
QY 327 KDDSQFSGSSEC-----TERPPCTTKDYFQIHPTPCDEEGKTQIMYKWIIEPKICRE----- 376  
Db 1466 PDNDNY--YKKCPIHNKTEIEKRT--DFETINSLCLCEEGYEPINMNKIKDTSRDYHYNN 1521  
QY 377 -----DLTDAIRLPPSGEKKDCPPCNPFGFYNNSSSS--CHPCPPG--TFSDGTK---E 422  
Db 1522 IFVTYKNLSYIV-----NDNNICMECNLGFYKNNISSEKCIKPRSTTTTYGSKSIQN 1576  
QY 423 CRPCPAGTEPALGFYKWNVLPGNMKTSC-----FNVGNS-KCDGMNGWEVAGDHI- 473  
Db 1577 CNSCHKG-----YK-----DQKQVCSKCLPNHF'CVGKSQKNDKNNLSQYAGDAII 1622  
QY 474 -----QSGAGGSDNDYLILNLH-----IPG 493  
Db 1623 CPNYSVTLPYENNISFKNCLCIKYEKNFQDFYNNHNCKKAPLNFYKDTISNDLSIP- 1681  
QY 494 FKPTSTMTGATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVESWGCTKEKQAYTH 553

Db 1682 -CPTNSITLNTGA-----TSIYNICIDKG---FFDYTVSYSCVNCPHGY-YCSEKDMT7K 1731

QY 554 I-----IFKNATFTTFAQRTNQGDNRFRFNDVMKIYSITATNAVDGVASSCRACALG 608

Db 1732 LAPPIKCPNRYTIIYKGSYNISHCVESGYTVNTVIVEHS-SENGIIESHESAKNIKLO 1790

QY 609 SEQSGSSCVPCPPGHYIEKETN-QCKECPDPT-----YLSIHQVYGKEACI 653

Db 1791 AKNRTSVCKVQPOSSYKSKISNEQCHCKPNKSKTLKDFNNSDIFFLCTMGYYTDKKECK 1850

QY 654 PCG-----PGSKNNQDHSVYSCDFFYHEKENQILHYDFSNLSSVGSMLMNGPSFTSKGT 707

Db 1851 PCWFNKLYCEGEKTYQIETIIYDEIILNQYIRLLSDISTKKSFINTIIN--EYINKGL 1908

QY 708 KYFHFFNISLCGHEGKKMALCTNNITDFTVKEIVAGSDDYT-----NLVGAFV-----CQS 758

Db 1909 -ISEITHISMMEKSRKSG--NNIQ--SIRKAKKDSEIYAVNIYRNLGNIITKDRELKT 1952

QY 759 TIIPSESKGFRAALSSOSIILADTFIGTVETTL---KNINIKEDMFPVPTSOIPDVHFF 815

Db 1963 PKINNQIKKIELAKSIENKIIAN---INSKLTLLSEGNKNIDTD-----TSIL--ANLF 2011

QY 816 YKSS-----TATTSCIN--GRSTAVKMRNPTKSGAGVISVPS---KCPAGTCDGCTEYF 865

Db 2012 YKSSENIYIKHQKLINCQNTVIPL-----GVDSSQNFDDCKCKKG-----YY 2055

QY 866 L-----WESAEACPLCTEHDFFEIEG-----ACKRGFQETL 896

Db 2056 LEDRIVLKNIKICKPCPEGTFKNYGDGVKKCISCPPKSTSIKSIYPNHCFCKNGF---F 2112

QY 897 YVWNEPKWCIKGISLPEKKLATC 919

Db 2113 YSKDTCLELEG-----ATC 2127

RESULT 14

Q99ND0

ID Q99ND0 PRELIMINARY; PRT; 5374 AA.

AC Q99ND0;

DT 01-JUN-2001 (Tremblrel. 17, Created)

DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE ZAN (Zonadhesin).

GN ZAN.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=129/SV;

RX MEDLINE=21138439; PubMed=11239002;

RA Wilson M.D., Riemer C., Martindale D.W., Schnupf P., Boright A.P.,

RA Cheung T.L., Hardy D.M., Schwartz S., Scherer S.W., Tsui L.-C.,

RA Miller W., Koop B.F.;

RT "Comparative analysis of the gene-dense ACHE/TFR2 region on human

RT chromosome 7q22 with the orthologous region on mouse chromosome 5.,";

RL Nucleic Acids Res. 29:1352-1365(2001).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=129/SV;

RA Cheung T.L., Wilson M.D., Koop B.F., Hardy D.M.;

RT "Genomic Basis of Inter- and Intra-species Variation in Zonadhesin

RT Domain Structure.,";

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 3 MAM DOMAINS.

DR EMBL: AF312033; AAK28824.1; -.

DR EMBL: AY046056; AAL04416.1; -.

DR MGI:106656; Zan.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR003645; FOLN.

DR InterPro: IPR000998; MAM\_domain.

DR InterPro: IPR003328; Tila\_Cysrich.

DR InterPro: IPR002919; TIL\_Cysrich.

DR InterPro: IPR001007; VWF\_C.

DR InterPro: IPR001846; VWF\_D.

DR Pfam: PF00629; MAM; 3.

DR Pfam: PF01826; TIL; 25.

DR Pfam: PF02345; Tila; 25.

DR Pfam: PF00094; vwd; 4.

DR SMART: SM00001; EGF\_like; 1.

DR SMART: SM00274; FOLN; 21.

DR SMART: SM00137; MAM; 3.

DR SMART: SM00214; VWD; 25.

DR SMART: SM00216; VWD; 4.

DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.

DR PROSITE: PS01186; EGF\_2; 18.

DR PROSITE: PS50060; MAM\_2; 3.

DR EGF-like domain; Glycoprotein.

QW EGF-like domain; Glycoprotein.

SQ SEQUENCE 5374 AA; 579536 MW; 90D2D8CFE5DE24EB CRC64;

Query Match 3.4%; Score 193.5; DB 11; Length 5374;

Best Local Similarity 19.3%; Pred. No. 2.8e-06;

Matches 233; Conservative 103; Mismatches 387; Indels 487; Gaps 64;

QY 23 RGRSPWPSPAWI-----CCWALAGCQ-----AAWAGDLPSSSR-----PLPPCQE 63

Db 3122 OGALIEGKWTWITSGCTQSCNCTGGAIQCQNFQCPKLTCKDLKDGSSNCTNIPL-----Q 3177

QY 64 KDYHFYEYTECDSSGSRWRVAIPNSAVDCSGL---PDPVRGKECTFSC-ASGEYLEMKNQ- 118

Db 3178 CPAHSRYTNCPLPS-----CPPSCLDPEGLCEGTSPKVPSTCREGCICQPGYLMHKNC 3230

QY 119 -----VCS-----KCGEGTYSLSGSGIKFDE 138

Db 3231 VLRIFCGCKNTQGAFTSADKTWISRGCTQSCTSAGAIHCNFKCPSGTCKNGDNGSSN 3290

QY 139 WDELPAAGFSNIATFMDTVVGPDSRDPGCGNNSSWIPRGNYIESNRDDCTVSLIYAVHLKK 198

Db 3291 CTEITLQCPTNSQFTDCL---PSCVPSCSNRCEVTSPSPSSCREGC-----LCN 3337

QY 199 SGYVFFEYQYVDNNIFFEIFFIONDQCQEMDTTTTDKWVKLTNDNGEWGSHVMLKSGTNILY 258

Db 3338 HGFVFSE-----DKCVPRTQCGCKDARGAIIPAG----- 3366

QY 259 WRTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCPKCTFSPKPGSFNCQ--VCIPRNTYS 316

Db 3367 -----KWTWSKGTQSCACV-----EGNIQCQNFQCCPPETY- 3397

QY 317 EKGAKECIRCKDDSQFSGSSECTE---RPPCTTKDYFQIHTPC-----DEEGKTQI 364

Db 3398 -----CKDNSE--GSSTCTKITLQCPAHTQ-----YTSCPLSCLPCLDPEG----- 3437

QY 365 MYKWIEPKI---CRE-----DLTDAL--RLPPSGEKKDCPPCNP 399

Db 3438 LCKDISPKVPSTCKEKGVCQSGYVLSNCKVLAECDCCKDAQALIPAGKTWTSPGCTQS 3497

QY 400 -FYNNGSSSCH--PCPPGTF---SDGTEKCR---PCPA----- 428

Db 3498 CACMGGAQCQSSQCPGTYCKDNEGNSNCAKITLQCPAHSLETNCLPCLPCLDPPDG 3557

QY 429 --GTEPALGFYKWNVLPGNMKTSC-----FNVGNSKCDGMN----- 464

Db 3558 LCKGASPK-----VPSTCKEKGTCQSGYVLSNCKLLRNRCCKDAHALIPEDK 3607

QY 465 -----GWEV-----AGDHIQSGAGSDNDYLI-----LNLHI----- 491

Db 3608 TWVSRGCTQSCVCTGGSIQCLSFQCQPPGAYCKDNEGSSNCARIPPOCPANSHYTDCTFP 3667

QY 492 --PGFKPPTSMTCATGSELGRITFVFETLCSADCVLY--FMVDINRKSTNVVSWGGTKE 547

Db 3668 CPPSCSDPEGHCEASGPRV-----PSTCRGCLCNPGFVLD-RDKCVPRVEC--GCKD 3717

QY 548 KQAYTHIIFKKNATFTTFAFQ-----RTNQ-----GQDNRREFINDMKVI--YS 588

Db 3718 AQG--ALIPSGKWTWSPGRTQSCACMGVVCQSSQCPGTYCKDNEGNSNCAKITLQ 3775







GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 00:12:42 ; Search time 4334.16 Seconds  
(without alignments)  
13287.722 Million cell updates/sec

Title: US-10-073-333A-3  
Perfect score: 3556  
Sequence: 1 atgctgttcgcgcgcgggg.....ggtaaaaaaaaaaaaaaa 3556

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_estl1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
C 1	704	19.8	705	14	BQ182610
2	660.4	18.6	697	14	BM723980
3	654.4	18.4	656	14	BM783507
C 4	651.2	18.3	713	14	BQ772508
5	624.8	17.6	930	12	BF035563
6	610.4	17.2	646	10	AW954806

7	584.4	16.4	896	12	BG116864
8	553.2	15.6	762	13	BI227132
9	527	14.8	527	14	BM752466
10	513.4	14.4	667	10	BB384282
11	512.8	14.4	634	14	BQ551364
12	507.4	14.3	515	14	BM771726
C 13	504	14.2	576	14	W60844
14	498.4	14.0	625	12	BG724280
15	494.2	13.9	801	13	BI731892
16	493	13.9	541	10	BE237235
17	483	13.6	483	14	BM759885
18	482.2	13.6	557	14	W60843
19	481.4	13.5	484	9	AA010992
C 20	478	13.4	478	10	AW665873
C 21	477	13.4	477	9	AI885785
C 22	472.6	13.3	585	14	BM974911
23	470	13.2	1773	11	AK017241
24	468.8	13.2	718	12	BG086402
25	449.4	12.6	508	13	BM538241
C 26	446	12.5	458	12	BF740039
27	435.6	12.2	1109	14	BQ896007
28	424.8	11.9	600	10	BB617824
29	421.8	11.9	656	10	BB626387
30	419.2	11.8	653	10	BB650824
C 31	414.4	11.7	507	9	AI797353
32	412.4	11.6	565	13	BI337511
33	404	11.4	420	13	BI480715
C 34	397	11.2	445	9	AA010869
35	394.4	11.1	641	10	BB629056
36	391.8	11.0	667	13	BM491171
37	388.6	10.9	640	10	BB622295
38	381.8	10.7	859	14	BQ730700
39	374.6	10.5	449	14	H18553
40	374	10.5	374	9	AA644099
41	370.8	10.4	446	9	AI317059
C 42	369.4	10.4	395	10	AW444960
43	344.4	9.7	553	10	BB644517
44	343.8	9.7	631	10	AV335504
C 45	338	9.5	603	13	BJ081522

ALIGNMENTS

RESULT 1  
BQ182610/C  
LOCUS BQ182610 705 bp mRNA linear EST 30-APR-2002  
DEFINITION UI-H-EU0-azx-a-23-0-UI.s1 NCI\_CGAP\_Carl Homo sapiens CDNA clone  
IMAGE: 5854750 3', mRNA sequence.  
ACCESSION BQ182610  
VERSION BQ182610.1 GI:20358160  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 705)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Dr. Jose Mercuende  
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
The following repetitive elements were found in this CDNA sequence: 1-40, >AT-rich#Low\_complexity (matched complement)  
Seq primer: M13 FORWARD  
POLYA=Yes.





```
QY 1835 CGGGTTCATCGTGTGTCCCTGCCCTCAGGCCACTACAT-TGAGAAAGAAACCAACCAG 1893
Db 1 CGGGTTCATCGTGTGTCCCTGCCCTCAGGCCACTACATNTGAGAAAGAAACCAACCAG 60

QY 1894 TGCAAGGAATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGAGGCT 1953
Db 61 TGCAAGGAATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGAGGCT 120

QY 1954 TGTATTCCATCGGGGCCTGGGAGTAAACAATCAGGACCATTCGGTTTGCTATAGTGAC 2013
Db 121 TGTATTCCATCGGGGCCTGGGAGTAAACAATCAGGACCATTCGGTTTGCTATAGTGAC 180

QY 2014 TGCTTTTCTACCATGAAAAAGAAATCAGATTTTGCACATATGACTTTAGCAACCTCAGC 2073
Db 181 TGCTTTTCTACCATGAAAAAGAAATCAGAGTTTGCACTATGACTTTAGCAACCTCAGC 240

QY 2074 AGTGTGGGCTCATTAATGAATGGCCCGAGCTTCACCTCCAAAGGAACAAAATACTTCCAT 2133
Db 241 AGTGTGGGCTCATTAATGAATGGCCCGAGCTTCACCTCCAAAGGAACAAAATACTTCCAT 300

QY 2134 TTCTTCAATATCAGTTTATGTGGGCATGAGGGGAAGAAGATGCTCTCTGTACCAACAAT 2193
Db 301 TTCTTCAATATCAGTTTATGTGGGCATGAGGGGAAGAAGATGCTCTCTGTACCAACAAT 360

QY 2194 ATAACAGACTTTACAGTAAAGAAATAGTGGCAGGGTCAGATGATTAACACAAA-TTTGGT 2252
Db 361 ATAACAGACTTTACAGTAAAGAAATAGTGGCAGGGTCAGATGATTAACACAAAATTTTGGT 420

QY 2253 AGGGCATTTGTATGCCAGTCAACAATATTCTTCTGAAAGTAAGGGTTTCCGAGCAGC 2312
Db 421 AGGGCATTTGTATGCCAGTCAACAATATTCTTCTGAAAGTAAGGGTTTCCGAGCAGC 480

QY 2313 CTTATCATCAAAATCCATCATTTCTGGCAGATACATTCATAGGAGTCACAGTTGAAACCCAC 2372
Db 481 CTTATCATCAAAATCCATCATTTCTGGCAGATACATTCATAGGAGTCACAGTTGAAACCCAC 540

QY 2373 ATTGAAAAATATTAATATAAAAGAAGATATGTTCCCAGTTCCAACAAGCCAAATACCCAGA 2432
Db 541 ATTGAAAAATATTAATATAAAAGAAGATATGTTCCCAGTTCCAACAAGCCAAATACCCAGA 600

QY 2433 TGTGCATTTCTTTTATAAGTCTTCTACAGCAACAACATCTTGTTAATGGCCGATCAAC 2492
Db 601 TGTGCATTTCTTTTATAAGTCTTCTACAGCAA-AACATCTTGTTAATGGCCGATCAAC 659

QY 2493 TGCTGTGAAAATGAGGTGTAATCCTACTAAATCTGGAG 2530
Db 660 TGCTGTGAAAATGAGGTGTAATCCTACTAAATCTGGAG 697
```

```
RESULT 3
BM783507
LOCUS
DEFINITION BM783507 656 bp mRNA linear EST 05-MAR-2002
K-EST0061445 S7SNU719 Homo sapiens cDNA clone S7SNU719-36-H08 5',
mRNA sequence.
ACCESSION BM783507
VERSION BM783507.1 GI:19131739
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 656)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
```

```
FEATURES
source
1. 656
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S7SNU719-36-H08"
/clone_lib="S7SNU719"
/sex="M"
/tissue_type="Stomach"
/cell_type="Epithelial"
/cell_line="SNU-719"
/lab_host="Top10p"
/note="Organ: Stomach; Vector: pCNS; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transfection of
competent cells E. coli Top10p, by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."
BASE COUNT 191 a 147 c 145 g 173 t
ORIGIN
```

```
Query Match 18.4%; Score 654.4; DB 14; Length 656;
Best Local Similarity 99.8%; Pred. No. 1.3e-121;
Matches 655; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1710 TAATCAGGGTCAAGATAATAGACGGTTCATCAATGACATGGTGAAGATTTATCTATCAC 1769
Db 1 TAATCAGGGTCAAGATAATAGACGGTTCATCAATGACATGGTGAAGATTTATCTATCAC 60

QY 1770 AGCCACTAATGCAGTTGATGGGTGGCGTCTCATGCCGTGCCTCGCTCGCTTCTGA 1829
Db 61 AGCCACTAATGCAGTTGATGGGTGGCGTCTCATGCCGTGCCTCGCTCGCTTCTGA 120

QY 1830 ACAGTCGGGTTTCATCGTGTGCCCTGCCCTCCAGGCCACTACATGAGAAAGAAACCAA 1889
Db 121 ACAGTCGGGTTTCATCGTGTGCCCTGCCCTCCAGGCCACTACATGAGAAAGAAACCAA 180

QY 1890 CCAGTGAAGGAATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGA 1949
Db 181 CCAGTGAAGGAATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGA 240

QY 1950 GGCTTGATTCCATGCGGGCCTGGGAGTAAACAATCAGGACCATTCGGTTTGCTATAG 2009
Db 241 GGCTTGATTCCATGCGGGCCTGGGAGTAAACAATCAGGACCATTCGGTTTGCTATAG 300

QY 2010 TGACTGCTTTTCTACCATGAAAAAGAAATCAGATTTTGGCACTATGACTTTAGCAACCT 2069
Db 301 TGACTGCTTTTCTACCATGAAAAAGAAATCAGAGTTTGGCACTATGACTTTAGCAACCT 360

QY 2070 CACGAGTGTGGGCTCATTATGAATGGCCCCCAGCTTCACCTCCAAAGGAACAAAATACTT 2129
Db 361 CACGAGTGTGGGCTCATTATGAATGGCCCCCAGCTTCACCTCCAAAGGAACAAAATACTT 420

QY 2130 CCATTTCTTCAATATCAGTTTATGTGGGCATGAGGGGAAGAGATGGCTCTCTGTACCAA 2189
Db 421 CCATTTCTTCAATATCAGTTTATGTGGGCATGAGGGGAAGAGATGGCTCTCTGTACCAA 480

QY 2190 CAATATAACAGACTTTTACAGTAAAAAGAAATAGTGGCAGGTCAGATGATTACACAAATTT 2249
Db 481 CAATATAACAGACTTTTACAGTAAAAAGAAATAGTGGCAGGTCAGATGATTACACAAATTT 540
```

Email: yongsung@mail.kribb.re.kr  
Plate: 36 row: H column: 08  
High quality sequence stop: 656.

QY	2250	GGTAGGGGCA	TTGTATGCCAGTCAACA	ATTATTCCTTCTGAAAGTAAGGGTTTCCGAGC	2309
Db	541	GGTAGGGGCA	TTGTATGCCAGTCAACA	ATTATTCCTTCTGAAAGTAAGGGTTTCCGAGC	600
QY	2310	AGCCTTATCATCA	CAATCCCATCATCTTGGCAGATACATTCATAGGAGTCACAGTTG	2365'	
Db	601	AGCCTTATCATCA	CAATCCCATCATCTTGGCAGATACATTCATAGGAGTCACAGTTG	656	
RESULT 4					
BQ772508/c					
LOCUS			713 bp	mRNA	linear
DEFINITION			UI-H-FEO-bbq-b-01-0-UI.s1	NCI_CGAP_FEO	Homo sapiens
ACCESSION			BQ772508		
VERSION			BQ772508.1	GI:21980984	
KEYWORDS			EST.		
SOURCE			human.		
ORGANISM			Homo sapiens		
REFERENCE			1 (bases 1 to 713)		
AUTHORS			NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.	
TITLE			National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL			Unpublished (1997)		
COMMENT			Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: James Martin CDNA Library preparation: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bentso-soares@uiowa.edu The following repetitive elements were found in this cDNA sequence: 1-40, >AT-rich#Low_complexity (matched complement) Seq primer: M13 FORWARD POLYA=Yes.		
FEATURES			Location/Qualifiers		
source			1..713		
			/organism="Homo sapiens"		
			/db_xref="taxon:9606"		
			/clone="UI-H-FEO-bbq-b-01-0-UI"		
			/clone_lib="NCI_CGAP_FEO"		
			/tissue_type="Chondrosarcoma Cell line"		
			/dev_stage="Adult"		
			/lab_host="DH10B (Life Technologies)"		
			/note="Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; NCI_CGAP_FEO is a cDNA library containing the following tissue(s): a pool of 3 chondrosarcoma cell lines ( grade 2) The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGTACGGAC. The cell lines was provided by Dr James Martin of University of Iowa.		
			TAG_LIB=UI-H-FEO		
			TAG_TISSUE=Human grade 2 chondrosarcoma cell line pool		
			TAG_SEQ=CGCTACGGAC"		
BASE COUNT	193 a	136 c	120 g	258 t	6 others
ORIGIN					
Query Match			18.3%	Score 651.2;	DB 14; Length 713;
Best Local Similarity			98.2%	Pred. No. 5.7e-121;	
Matches 699;			Conservative 0;	Mismatches 7;	Indels 6; Gaps 4;
QY	2851	AATCAAAAAC	TGGATACAAATATTC	CAAGTAGTAGTAATGACGACTAACTCA-AAAGAGTG	2909

Db	713	AATCAANA	ACTGGAATACAAATATNCCAAGTTAGTAATGACGACTAACTCANAAAGAGTG	654
QY	2910	TGAAC	TCCCGCTGCAGACAGTTTGTCTATCATGGAAGGAGAAGATAATGAAGAGGAAGT	2969
Db	653	TGAAC	TCCCGCTGCAGACAGTTTGTCTATCATGGAAGGAGAAGATAATGAAGAGGAAGT	594
QY	2970	TGTATAT	TCCAATAAACAAGTCACTACTAGGAAAACCTCAAACTCTTTGGCAACCAAGGAAAA	3029
Db	593	TGTATAT	TCCAATAAACAAGTCACTACTAGGAAAACCTCAAACTCTTTGGCAACCAAGGAAAA	534
QY	3030	AGAAG	ACCATTTTGAATCTGTTCAACTGAAAACCTCAAGATCCCCAAATATATGAAGAGA	3089
Db	533	AGAAG	ACCATTTTGAATCTGTTCAACTGAAAACCTCAAGATCCCCAAATATATGAAGAGA	474
QY	3090	CAGT	GTGTAGCCTTGAGACTAATGAACAAGAAAACCTGCTCTAGTTTACAGGACCATA	3149
Db	473	CAGT	GTGTAGCCTTGAGACTAATGAACAAGAAAACCTGCTCTAGTTTACAGGACCATA	414
QY	3150	TTTT	AGGGTCTGCTCATACCTGTGCACATTTGGTGATCTCACAGAGGAGGGCCCATGCCGC	3209
Db	413	TTTT	AGGGTCTGCTCATACCTGTGCACATTTGGTGATCTCACAGAGGAGGGCCCATGCCGC	354
QY	3210	TGAAA	AGGGAAGGAGATTGAAACATTTTGATTGCTTATCACATGGTCAAGTACCTTGCCA	3269
Db	353	TGAAA	AGGGAAGGAGATTGAAACATTTTGATTGCTTATCACATGGTCAAGTACCTTGCCA	294
QY	3270	AATAA	AGGAAAGCAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAAGAGA	3329
Db	293	AATAA	AGGAAAGCAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAAGAGA	234
QY	3330	TTTAT	CTGTATATACATAACTGAAAACCAAGTTTAAGCCCCACCAATGCACCTGCTGATG	3389
Db	233	TTTAT	CTGTATATACATAACTGAAAACCAAGTTTAAGCCCCACCAATGCACCTGCTGATG	174
QY	3390	CATG	CCATATAATTAATGGTAACTTTTATTTCTTTATGATGTCTACATAA-CAAGTGTGA	3448
Db	173	CATG	CCATATAATTAATGGTAACTTTTATTTCTTTATGATGTCTACATAA-CAAGTGTGA	114
QY	3449	TTTG	GAAGGCACA-TGTGAGCATATGCATTAT--GATCCAATTTATGTTTTCTTTGT	3504
Db	113	TTTG	GAAGGCACANTGTGAGCATATGCATTNNATGATCCCAATTTATGTTTTCTTTGT	54
QY	3505	TTAT	ATTTTGGGAAAATTAATAATTTTAAAGGTAAATAAAAAAAAAAAAAA	3556
Db	53	TTAT	ATTTTGGGAAAATTAATAATTTTAAAGGTAAATAAAAAAAAAAAAAA	2

RESULT 5  
BF035563  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BF035563  
601457452F1 NIH\_MGC\_66 Homo sapiens cDNA clone IMAGE:3861156 5',  
mRNA sequence.  
BF035563  
BF035563.1 GI:10743303  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 930)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTP  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>





```
|||||
Db 481 CCTGTCCATACATCAGGTCTATGGCAAAGAGCGCTTGATTCCATGCGGCGCTGGGAGTAA 540
QY 1980 AAACAATCAGGACCAATTCGGTTTGCTATAGTACTGCTTTTCTACCATG-AAAAAGAAA 2038
Db 541 AAACAATCAGGACCCCTCGGTTTGCTATAGTACTGCTTTTCTACCATGAAAAAGAAA 600
QY 2039 ATCAGATTTTGCACATATGAC-TTTAGCAACCTCAGCAGTGTGGG 2081
Db 601 ATCAGAGTTTGCACATATGACNTTTGGCAACCTTAACAGGGTGGG 644

RESULT 7
BG116864
LOCUS
DEFINITION 602318965F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4423217 5',
            mRNA sequence.
ACCESSION BG116864
VERSION BG116864.1 GI:12610303
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE 1 (bases 1 to 896)
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-re@mail.nih.gov
          Tissue Procurement: ATCC
          CDNA Library Preparation: Life Technologies, Inc.
          CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM10165 row: b column: 18
          High quality sequence stop: 740.
          Location/Qualifiers
            1..896
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:4423217"
              /clone_lib="NIH_MGC_88"
              /tissue_type="duodenal adenocarcinoma, cell line"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: small intestine; Vector: pCMV-SPORT6;
              Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
              oligo-dT primed. Average insert size 1.767 kb. Library
              enriched for full-length clones and constructed by Life
              Technologies. Note: this is a NIH_MGC Library."
              299 a 145 c 214 g 238 t

BASE COUNT 299 a 145 c 214 g 238 t
ORIGIN

Query Match 16.4%; Score 584.4; DB 12; Length 896;
Best Local Similarity 97.4%; Pred. No. 1.5e-107;
Matches 637; Conservative 0; Mismatches 11; Indels 6; Gaps 4;

QY 2889 GACGACTAACTCAAAAGAGTGTGAACTCCCGGCTGCAGACAGTGTGCTATCATGGAAGG 2948
Db 1 GACGACTAACTCAAAAGAGTGTGAACTCCCGGCTGCAGACAGTGTGCTATCATGGAAGG 60
QY 2949 AGAAGATATGAAGAGGGAAGTTGTATATTCCAAATAAACAGTCACACTAGGAAAACTCAA 3008
Db 61 AGAAGATATGAAGAGGGAAGTTGTATATTCCAAATAAACAGTCACACTAGGAAAACTCAA 120
QY 3009 ATCTTTGGCACCACCAAGGAAAAAGAGACCATTTTGAATCTGTTCACTGAAACCTCAAG 3068
Db 121 ATC-TTGGCACCACCAAGGAAAAAGAGACCA-TTTGAATCTGTTCACTGAAACCTCAAG 178
QY 3069 ATCCCCAAATATATGAAGAGACAGTGTGTAGCCTTGAGACTAATGAACAAAGAAACCTG 3128
|||||
```

```
Db 179 ATCCCCAAATATATGAAGAGACAGTGTGTAGCCTTGAGACTAATGAACAAAGAAACCTG 238
QY 3129 CTCTAGTTTTACAGGACCATATTTTAGGGTCTGTCTCCTCATACCTGTCACATTGGTGATCT 3188
Db 239 CTCTAGTTTTACAGGACCATATTTTAGGGTCTGTCTCCTCATACCTGTCACATTGGTGATCT 298
QY 3189 CACAGAGGAGGGCCATGCCGCTGAAAAGGGAGGAGAGATTGAACATTTGATGGCTTATC 3248
Db 299 CACAGAGGAGGGCCATGCCGCTGAAAAGGGAGGAGAGATTGAACATTTGATGGCTTATC 358
QY 3249 ACATGGTCAAGTACCTTGCCAAATAAAGGAAAGCAAAATGATTTGGGTCTCAACTGAAGAT 3308
Db 359 ACATGGTCAAGTACCTTGCCAAATAAAGGAAAGCAAAATGATTTGGGTCTCAACTGAAGAT 418
QY 3309 GAAGCTCAACTCAGGAAGAGAGATTATCTGTATATACACATAACTGAAAACCAAGTTTAAG 3368
Db 419 GAAGCTCAACTCAGGAAGAGAGATTATCTGTATATACACATAACTGAAAACCAAGTTTAAG 478
QY 3369 CCCACCAATGCACCTGCTGATGCATGCCATATATAATTAATGGGTAACTTTTATCTTTATGA 3428
Db 479 CCCACCAATGCACCTGCTGATGCATGCCATATATAATTAATGGGTAACTTTTATCTTTATGA 538
QY 3429 TGTCTACATAACAAAGTGTGATTTGGAAGGCACATGTGAGCATATGCATTAATGATCCAATT 3488
Db 539 TGTCTACATAACAAAGTGTGATT--GGAAGGACATGTGAGCATATGCATTAATGATCCAATT 596
QY 3489 TATGTTTTTTCTTTGTATATATTTTGGGGAAAAATTTAAATTTTAAAGGTAAA 3542
Db 597 TATGTTTTTTCTT--GTTTATATTTTGGGGAAAAATTAACATTTGTTTAAAGGTATA 648

RESULT 8
BI227132
LOCUS
DEFINITION 602952315F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:5096643 5',
            mRNA sequence.
ACCESSION BI227132
VERSION BI227132.1 GI:14680576
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE 1 (bases 1 to 762)
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-re@mail.nih.gov
          Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
          CDNA Library Preparation: Ling Hong/Rubin Laboratory
          CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM1868 row: n column: 04
          High quality sequence stop: 659.
          Location/Qualifiers
            1..762
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:5096643"
              /clone_lib="NIH_MGC_8"
              /tissue_type="Burkitt lymphoma"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
              EcoRI; CDNA made by oligo-dT priming. Directionally
              cloned into EcoRI/XhoI sites using the following 5',
              adaptor: GGCACGAG(G). Size-selected >500bp for average
              insert size 1.8kb. Library constructed by Ling Hong in
              the laboratory of Gerald M. Rubin (University of
              California, Berkeley) using ZAP-cDNA synthesis kit
```

```

BASE COUNT      225 a   176 c   178 g   183 t
ORIGIN
Query Match      15.6%; Score 553.2; DB 13; Length 762;
Best Local Similarity 96.2%; Pred. No. 3.1e-101;
Matches 612; Conservative 0; Mismatches 13; Indels 11; Gaps 4;

QY 890 CATTGAGCAACAAACCAGGTTTCATTCAACTGCCAGGTGTGCCAGAAACACCTATTCTG 949
Db 2 CATTGAGCAACAAACCAGGTTTCATTCAACTGCCAGGTGTGCCAGAAACACCTATTCTG 61

QY 950 AGAAAGGAGCCAAAGAATGTATAGGTGTAAAGACGACTCTCAATTTTC-----AGGAT 1003
Db 62 AGAAAGGAGCCAAAGAATGTATAGGTGTAAAGACGACTCTCAATTTTCAGAGGAAGGAT 121

QY 1004 CCAGTGAGGTACAGAGCGCCCTCCCTGTACCACAAAGACTATTTCCAGATCCATACTC 1063
Db 122 CCAGTGAGGTACAGAGCGCCCTCCCTGTACCACAAAGACTATTTCCAGATCCATACTC 181

QY 1064 CATGTGATGAAGAAGGAAAGACACAGATAAATGTACAAGTGGAGAGCCCAAAATCTGCC 1123
Db 182 CATGTGATGAAGAAGGAAAGACACAGATAAATGTACAAGTGGAGAGCCCAAAATCTGCC 241

QY 1124 GGGAGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAAGAGGATTGTCCGC 1183
Db 242 GGGAGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAAGAGGATTGTCCGC 301

QY 1184 CTTGCAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCCCTCGGAA 1243
Db 302 CTTGCAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCCCTCGGAA 361

QY 1244 CATTTTCAGATGGAACCAAGAAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCAC TTG 1303
Db 362 CATTTTCAGATGGAACCAAGAAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCAC TTG 421

QY 1304 GCCTT-TGAATATAAATGGTGGATGTCTCCTCGGCAACATGAAAACTTCCTGCTTCAAT 1362
Db 422 GCCTTCTGAATATAAATGGTGGATGTCTCCTCGGCAACATGAAAACTTCCTGCTTCAAT 481

QY 1363 GTTGGGAATTCAAAGTCCGATGG-AATGAATGGTTGGGAGGTGGCTGGAGATCATATCCA 1421
Db 482 GTCCGGGAATTCAAAGTCCGATGGCAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCA 541

QY 1422 GAGTGGGCTGGAGGTTCTGACAATGATTACCTGATCTTAAACTTGCATATCCCAGG--- 1478
Db 542 GAGTGGGCTGGAGGTTCTGACAATGATTACCTGATCTTAAACTTGCATATCCCAGGCA 601

QY 1479 ATTTAAACCACCAACATCTATGACTGGAGCCACGGG 1514
Db 602 TCTAACCAGCAACATCTATGACTGGAGCCACCGG 637

RESULT 9
BM752466
LOCUS
DEFINITION K-EST0029003 S9SNU601 Homo sapiens cDNA clone S9SNU601-17-H12 5',
mRNA sequence.
ACCESSION BM752466
VERSION BM752466.1 GI:19082084
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 527)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
```

```

Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 17 row: H column: 12
High quality sequence stop: 527.

FEATURES
Location/Qualifiers
1..527
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S9SNU601-17-H12"
/clone_lib="S9SNU601"
/sex="M"
/tissue_type="Ascites"
/cell_type="Epithelial"
/cell_line="SNU-601"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pME18-FL3; Site_1: XhoI;
Site_2: XhoI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including SfiI
site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized with Superscript II using SfiI
oligo-dT primer. After first strand synthesis, RNA was
degraded by NaOH treatment and cDNA was amplified by PCR
reaction. The PCR products were digested with SfiI and
cloned into DraIII- digested pME18s-FL3 vector. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F, by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."
BASE COUNT      156 a   120 c   116 g   135 t
ORIGIN
Query Match      14.8%; Score 527; DB 14; Length 527;
Best Local Similarity 100.0%; Pred. No. 6.5e-96;
Matches 527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1461 AAAC TTGCATATCC CAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGTTCTGA 1520
Db 1 AAAC TTGCATATCC CAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGTTCTGA 60

QY 1521 ACTAGGAGAATAACATTTGTCTTTTGAGACCCCTCTGTTTCAGCTGACTGTGTTTGACTT 1580
Db 61 ACTAGGAGAATAACATTTGTCTTTTGAGACCCCTCTGTTTCAGCTGACTGTGTTTGACTT 120

QY 1581 CATGGTGGATATTAATAGAAAAAGTACAAATGTGGTAGAATCTGGGGTGGAAACCAAGA 1640
Db 121 CATGGTGGATATTAATAGAAAAAGTACAAATGTGGTAGAATCTGGGGTGGAAACCAAGA 180

QY 1641 AAAACAAGCTTACACCCCATATCATCTTCAAGAATGCAACTTTTACATTTACATGGGCATT 1700
Db 181 AAAACAAGCTTACACCCCATATCATCTTCAAGAATGCAACTTTTACATTTACATGGGCATT 240

QY 1701 CCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTTCATCAATGACATGGTGAAGATTTA 1760
Db 241 CCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTTCATCAATGACATGGTGAAGATTTA 300

QY 1761 TTCTATCACAGCCACTAATGCAGTTGATGGGTGGCGTCTCATGCCGTGCCTGTGCCCT 1820
Db 301 TTCTATCACAGCCACTAATGCAGTTGATGGGTGGCGTCTCATGCCGTGCCTGTGCCCT 360

QY 1821 CGGTTCTGAACAGTCGGGTTTCATCGTGTGCCCTGCCCTGCCCTCCAGGCCACTACATTGAGAA 1880
Db 361 CGGTTCTGAACAGTCGGGTTTCATCGTGTGCCCTGCCCTGCCCTCCAGGCCACTACATTGAGAA 420

QY 1881 AGAAACCAACCAGTGC AAGGAATGCCACCTGACACCTACCTGTCCATACATCAGGTCTA 1940
Db 421 AGAAACCAACCAGTGC AAGGAATGCCACCTGACACCTACCTGTCCATACATCAGGTCTA 480

QY 1941 TGGCAAAAGAGGCTTGTATTCCATCGCGGCGCTGGGAGTAAACAATC 1987
```



Db 481 TGGCAAAGAGGCTTGATTCATCGGGCCTGGGAGTAAAAACAATC 527  
|||||

RESULT 10  
BB384282  
LOCUS  
DEFINITION BB384282 RIKEN full-length enriched, 0 day neonate cerebellum Mus  
musculus cDNA clone C230033C19 3', mRNA sequence.  
ACCESSION BB384282  
VERSION BB384282.2 GI:16408254  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda  
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki  
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

TITLE  
JOURNAL  
COMMENT  
On Jul 13, 2000 this sequence version replaced gi:9104048.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa  
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and  
Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for  
further details.  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

FEATURES  
Location/Qualifiers  
1..667  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="C230033C19"  
/clone\_lib="RIKEN full-length enriched, 0 day neonate  
cerebellum"  
/tissue\_type="cerebellum"

/dev\_stage="0 day neonate"  
/lab\_host="DH10B"  
/note="Site 1: Sali; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGAAGGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 20.0 and subtraction to Rot = 479.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGAGATTCGAGTTAATTAAATTAATCCCCCCCCCCCC  
3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
modified pBluescript KS(+) after bulk excision from Lambda  
FLC I."

BASE COUNT 169 a 136 c 167 g 194 t 1 others  
ORIGIN

Query Match 14.4%; Score 513.4; DB 10; Length 667;  
Best Local Similarity 86.1%; Pred. No. 3.4e-93;  
Matches 568; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 2185 ACCAACANTATAACAGACTTTTACAGTAAAGAAATAGTGGCAGGGTCAGATGATTACACA 2244  
|||||

Db 8 ACCAACAAATATTTACAGACTNTACGGTAAAGAAATGGTGACAGGGTCAGATGATTACACA 67

QY 2245 AATTGGTAGGGGCATTTGTATGCCAGTCAACAATTTATTCCTCTGAAAGTAAGGGTTTC 2304  
|||||

Db 68 AATTGGTTGGAGCATTTGTATGCCAGTCAACTATATTTCTCTGAAAGTAAGGGCTTC 127

QY 2305 CGAGCAGCCTTATCANTCACAATCCATCATTTCTGGCAGATACATTCATAGGAGTCACAGTT 2364  
|||||

Db 128 CGGCAGCCCTGTCGTCACAGTCTATCATTTCTGGCAGATATGTTTATAGGTGTCACTGTT 187

QY 2365 GAAACCACATTTGAAAATATTAATATAAAGAAAGATATGTTCCAGTTCACAAAGCCAA 2424  
|||||

Db 188 GACACTGCCTTGCAAAACGTTAATATCAAGGAGGATATGTTCCAGTGTCAACCGAGCCAA 247

QY 2425 ATACCAGATGTGCATTTCTTTTATAAGTCTTCTACAGCAACACACATCTTGTATTAATGGC 2484  
|||||

Db 248 GTACCAGATGTTCAATTTCTTTTATAAGTCTTCCACAGCCACCACATCTTGTATTAATGGA 307

QY 2485 CGATCAACTGCTGTGAAAATGAGGTGTAATCCTACTACTAATCTGGAGCAGGAGTGATTTC 2544  
|||||

Db 308 CGATCGACTGCTGTGAAGATGAGGTGTAATCCCATGAGACCTGGTGAGGATGATCA 367

QY 2545 GTCCCAGCAAGTGCCCGCAGGATACCTGTGATGGGTGTACGTTCTATTTCTGTGGGAG 2604  
|||||

Db 368 GTCCCAGCAAGTGTCAGCTGGCACCTGTGATGGTGTACCTTCTACTTTCTATGGGAG 427

QY 2605 AGTGTCTGAAGCTTGCCCTCTGTGTGACGGAGCATGACTTCCATGAGATTGAGGGAGCCTGC 2664  
|||||

Db 428 AGTGCAAGAGCTTGCCCTCTGTGCACAGAACATGACTTCCATGAGATCGAGGAGCCTGC 487

QY 2665 AAGAGAGGATTTTCAGGAAACCTTCTATGTGTGGAATGAACCTAAATGGTGCATTAAGGA 2724  
|||||

Db 488 AAGAGAGGGCTTCAGGAAATATTTATATGTAATGAATGAACCTAAGTGGTGATTAAGGA 547

QY 2725 ATTTCTTTCCTTGAGAAAAAGTTGGCAACCTGTGAAACGGTTGACTTTGGCTGAAGGTG 2784  
|||||

Db 548 ATTTCTTTGCCCCGAAAGAGTTTCAACCTGCGAAACTGTTGACTTTTGGCTGAAGGTG 607

QY 2785 GGAGCCGGTGTGGGAGCTTTTACTGCCGTTTTTGTGGTGGCTCTGACCTGCTACTTCTGG 2844  
|||||

Db 608 GGAGCTGCTGTGGGCGCGTTTCACAGCCGTTTGTGGTGGCTTTAACATGCTACTTTTGG 667

RESULT 11  
BQ551364

LOCUS BQ551364 634 bp mRNA linear EST 20-JUN-2002  
DEFINITION H4008H01-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone  
H4008H01 5', mRNA sequence.  
ACCESSION BQ551364  
VERSION BQ551364.1 GI:21452250  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 634)  
AUTHORS VanBuren,V., Piao,Y., Dudekula,D.B., Qian,Y., Carter,M.G., Martin  
P.R., Stagg,C.A., Bassey,U., Aiba,K., Hamatani,T., Kargul,G.J.,  
Luo,A.G. and Ko,M.S.H.  
TITLE Assembly, verification, and initial annotation of NIA 7.4K mouse  
cDNA clone set  
JOURNAL Unpublished (2002)  
COMMENT Other\_ESTs: H4008H01-3  
Contact: Yong Qian  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA  
Email: cdna@lgsun.grc.nia.nih.gov  
This clone set has been freely distributed to the community. Please  
visit [http://lgsun.grc.nia.nih.gov/cDNA/NIA\\_7\\_4k.html](http://lgsun.grc.nia.nih.gov/cDNA/NIA_7_4k.html) for details.  
Plate: H4008 row: H column: 01  
Seq primer: -21M13 Reverse  
High quality sequence stop: 634  
POLYA-No. Location/Qualifiers  
1..634  
/organism="Mus musculus"  
/strain="C57BL/6"  
/db\_xref="niaEST:H4008H01-5"  
/db\_xref="taxon:10090"  
/clone="H4008H01"  
/clone\_lib="NIA Mouse 7.4K cDNA Clone Set"  
/sex="mixed"  
/dev\_stage="mixed"  
/lab\_host="DH10B"  
/note="Vector: pSPORT1; Site\_1: SalI; Site\_2: NotI; This  
clone is among a rearranged set of 7,407 clones from more  
than 20 cDNA libraries."  
BASE COUNT 180 a 143 c 165 g 146 t  
ORIGIN  
Query Match 14.4%; Score 512.8; DB 14; Length 634;  
Best Local Similarity 89.3%; Pred. No. 4.5e-93;  
Matches 566; Conservative 0; Mismatches 62; Indels 6; Gaps 1;  
QY 920 GCCAGTGTGTCCTCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAATGTATAAGGTGTA 979  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1 GCCAGATGTGTCCAGAAACACCTATTCTGAGAAAGGTCGTAAAGAATGTATTAGGTGTA 60  
QY 980 AAGCAGACTCTCAATTTTC-----AGGATCCAGTGAGTGTACAGAGCGCCCTCCCTGTA 1033  
||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 61 AGGAGACTCCCAAGTTTTCAGAGGAAGGAGCCAGTGAGTGTGGATCGCCCAACCTGCA 120  
QY 1034 CCACAAAGACTATTTCCAGATCCATACTCCATGTGATGAAGAAGGAAGACACAGATAA 1093  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 121 CTACCAAGACTATTTCCAGATCCATACCCCTGTGACGAAGAGGGAAGACACAGATAA 180  
QY 1094 TGTACAGTGGATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGATGCTATTAGATTGC 1153  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 181 TGTACAGTGGATAGAGCCCAAAATCTGCAGGAGGATCTCACAGATGCTATTAGACTGC 240  
QY 1154 CCCCTTCTGGAGAGAAGAGGATTGTCCGCCTTGCAACCCCTGGATTTTATAACAATGGAT 1213  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 241 CCCCTTCTGGAGAGAAGAGGATTGTCTCCTCTTGCAACCCAGGATCTATAACAATGGAT 300  
QY 1214 CATCTTCTGCCATCCCTGTCTCCTCGGAACATTTTCAGATGGAACCAAGAAATGTAGAC 1273  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 301 CGTCTTCTGCCATCCCTGCCCTCCGGGGACGTTTTTCAGATGGAACAAAGGAATGCAAGT 360

QY 1274 CATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAATGGTGAATGTCCTTC 1333  
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 361 CATGTCCGGCCGGAACAGAGCCAGCACTTGGGTTTGAATATAAATGGTGAATGTCCTTC 420  
QY 1334 CTGGCAACATGAAAACCTTCCTGCTTCAATGTTGGGAATTCAAAGTCGATGGAATGAATG 1393  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 421 CTGCCAACATGAAAACCTTCCTGCTTCAATGTTGGGAATTCCAAGTCGATGGAATGAACG 480  
QY 1394 GTTGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGTTCTGACAATGATTACC 1453  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 481 GTTGGAGGTGGCTGGAGATCATATCCGGAGTGGAGCTGGAGGCTCTGACAATGACTATC 540  
QY 1454 TGATCTTAAACTTGCATATCCACAGGATTTAAACCACCAACATCTATGACTGGAGCCACGG 1513  
| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 541 TCATCTTAAACTTGCACATCCACAGGATTTAAACCACCAACGTCATGACTGGAGCCACAG 600  
QY 1514 GTTCTGAAC TAGGAAGAATAACATTTGCTTTGA 1547  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 601 GTTCGAACTGGGAAGGATAACATTTGTGTTGA 634  
RESULT 12  
BM771726  
LOCUS  
DEFINITION K-EST0055671 S7SNU719s1 Homo sapiens cDNA clone S7SNU719s1-19-F04  
5', mRNA sequence.  
ACCESSION BM771726  
VERSION BM771726.1 GI:19101341  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 515)  
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished (2002)  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 19 row: F column: 04  
High quality sequence stop: 515.  
Location/Qualifiers  
1..515  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="S7SNU719s1-19-F04"  
/clone\_lib="S7SNU719s1"  
/sex="M"  
/tissue\_type="Stomach"  
/cell\_type="Epithelial"  
/cell\_line="SNU-719"  
/lab\_host="Top10P"  
/note="Organ: Stomach; Vector: pCNS; Site\_1: EcoRI;  
Site\_2: NotI; The poly (A)+ RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then decapped  
with tobacco acid pyrophosphatase (TAP). The decapped  
intact mRNA was ligated with DNA-RNA linker including EcoR  
I site by treatment of T4 RNA ligase and the first strand  
cDNA was synthesized from oligo dT-selected mRNA by  
priming with dT-tailed vector. The dT-tailed vector was  
adjusted to have about 60nt. The cDNA vector was  
circularized with E. coli DNA ligase after digestion of  
EcoRI which site is also included in vector. An RNA strand  
converted to a DNA strand by Okayama-Berg method. The  
obtained cDNA vectors were used for transformation of

competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library. After analyzing and sequencing about 2,000 - 3,000 colonies in original cDNA library, the abundant cDNAs were selected and amplified by PCR reaction using vector region primer including T7 promoter as 5' primer and N(dT)14 as 3' primer. The PCR products were used as template for synthesis of biotinylated single stranded RNA by in vitro transcription reaction. The synthesized RNA probes were hybridized with antisense single stranded cDNAs prepared from original library and incubated with avidin-gel. After removing DNA-RNA hybrids by centrifuge, the subtracted cDNA libraries were constructed by transformaion of the remaining DNA into competent cells E. coli Top10F' with electroporation method."

SE COUNT	150 a	119 c	114 g	132 t
IGIN				
Query Match	14.3%;	Score 507.4;	DB 14;	Length 515;
Best Local Similarity	99.8%;	Pred. No. 5.9e-92;		
Matches 508;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
1717	GGTCAAGATAATAGACGGTTTCATCAATGACATGGTGAAGATTTATTCTATCACAGCCACT	1776		
7	GGTCAAGATAATAGACGGTTTCATCAATGACATGGTGAAGATTTATTCTATCACAGCCACT	66		
1777	AATGCAGTTGATGGGTGGCGTCTCATGCGGTGCCTGTCCTCGGTTCTGAACAGTCG	1836		
67	AATGCAGTTGATGGGTGGCGTCTCATGCGGTGCCTGTCCTCGGTTCTGAACAGTCG	126		
1837	GGTTTCATCGTGTGTCCCTGCCCTCCAGGCCACTACATTGAGAAAGAAACCAACCAGTGC	1896		
127	GGTTTCATCGTGTGTCCCTGCCCTCCAGGCCACTACATTGAGAAAGAAACCAACCAGTGC	186		
1897	AAGGAATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGAGGGCTTGT	1956		
187	AAGGAATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGAGGGCTTGT	246		
1957	ATTCCATGCGGGCCTGGGAGTAAACAACATCAGGACCATTTCGGTTTGCTATAGTGACTGC	2016		
247	ATTCCATGCGGGCCTGGGAGTAAACAACATCAGGACCATTTCGGTTTGCTATAGTGACTGC	306		
2017	TTTTTCTACCATGAAAAGAAAATCAGATTTTGCACATATGACTTTAGCAACCTCAGCAGT	2076		
307	TTTTTCTACCATGAAAAGAAAATCAGAGTTTGCACATATGACTTTAGCAACCTCAGCAGT	366		
2077	GTGGGCTCATTAATGAATGGCCCCAGCTTCACCTCCAAAGGAACAAAATACTTCCATTTC	2136		
367	GTGGGCTCATTAATGAATGGCCCCAGCTTCACCTCCAAAGGAACAAAATACTTCCATTTC	426		
2137	TTCAATATCAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATA	2196		
427	TTCAATATCAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATA	486		
2197	ACAGACTTTACAGTAAAGAAAATAGTGGC	2225		
487	ACAGACTTTACAGTAAAGAAAATAGTGGC	515		
RESULT 13				
0844/c				
LOCUS	W60844	576 bp	mRNA	linear EST 15-OCT-1996
DEFINITION	zd26f10.s1 Soares_fetal_heart_NbHH19W	Homo sapiens cDNA clone		
	IMAGE:341803 3',	mRNA sequence.		
ACCESSION	W60844			
VERSION	W60844.1	GI:1367602		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 576)			

AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston ,R., Williamson,A., Wohldmann,P. and Wilson,R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1674 Std Error: 0.00  
Seq primer: mob.REGA+ET  
High quality sequence stop: 336.

FEATURES Location/Qualifiers  
1..576  
/organism="Homo sapiens"  
/db\_xref="GDB:1267178"  
/db\_xref="taxon:9606"  
/clone="IMAGE:341803"  
/clone\_lib="Soares\_fetal\_heart\_NbHH19W"  
/sex="unknown"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGCGCATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NbHL19W."  
BASE COUNT 171 a 112 c 108 g 182 t 3 others  
ORIGIN

Query Match	14.2%;	Score 504;	DB 14;	Length 576;
Best Local Similarity	94.4%;	Pred. No. 2.8e-91;		
Matches	543;	Conservative 0;	Mismatches 28;	Indels 4; Gaps 2;
Qy	2970	TGTATATTCCAATAAACAGTCACTACTAGGAAAACTCAAACTTTTGGCAACCAAGGAAAA	3029	
Db	575	TGTTTTTCCCATTAACCCGTCACTACTNGGAAACCTCAANTCTTTGGCAACCCAGGGAA	516	
Qy	3030	AGAAGA--CCATTTTGAATCTGTT-CAACTGAAAACCTCAAGATCCCCAAATATATGAA	3085	
Db	515	AAAGGAGACCCATTTTGAATCTGTTCACACTGAAACCTCAAGATCCCCAAATATATGAA	456	
Qy	3086	GAGACAGTGTGTAGCCCTTGAGACTAANTGAACNAAGAAACCTGCTCTAGTTTTACAGGAC	3145	
Db	455	GAGACAGTGTGTAGCCCTTGAGACTAANTGANCAAGAAACCTGCTCTAGTTTTACAGGAC	396	
Qy	3146	CATATTTTAGGGTCTGTCCCTCATACCTGTTCACATTTGGTGATCTCACAGAGGGGCCATG	3205	
Db	395	CATATTTTAGGGTCTGTCCCTCATACCTGTTCACATTTGGTGATCTCACAGAGGGGCCATG	336	
Qy	3206	CCGCTGAAAAGGGAAGGAGATTTGAAACATTTTGATTCCTTTATCACATGGTCAAGTACCTTT	3265	
Db	335	CCGCTGAAAAGGGAAGGAGATTTGAAACATTTTGATTCCTTTATCACATGGTCAAGTACCTTT	276	
Qy	3266	GCCAAATAAAGGAAAGCAAAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAA	3325	
Db	275	GCCAAATAAAGGAAAGCAAAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAA	216	
Qy	3326	GAGATTTATCTGTATATACACATACTGAAAACCAAGTTTAAAGCCCAACCAATGCACCTGCT	3385	
Db	215	GAGATTTATCTGTATATACACATACTGAAAACCAAGTTTAAAGCCCAACCAATGCACCTGCT	156	



QY	3386	GATGCATGCCATATAATTAATGGGTAAC	TTTATTCTTTATGATGCTACATAACAAGTG	3445
Db	155	GATGCATGCCATATAATTAATGGGTAAC	TTTATTCTTTATGATGCTACATAACAAGTG	96
QY	3446	TGATTTGGAAGGCACATGTGAGCATATG	CATATGATCCAAATTTATGTTTCTTTGTT	3505
Db	95	TGATTTGGAAGGCACATGTGAGCATATG	CATATGATCCAAATTTATGTTTCTTTGTT	36
QY	3506	TATATTTTGGGAAAAATAAAATTTT	TTTAAAGGTA	3540
Db	35	TATATTTTGGGAAAAATAAAATTTT	TTTAAAGGTA	1
RESULT 14				
LOCUS	BG724280	625 bp	mRNA	linear EST 08-MAY-2001
DEFINITION	602697903F1 NIH_MGC_97	Homo sapiens	CDNA clone	IMAGE:4829929 5', mRNA sequence.
ACCESSION	BG724280			
VERSION	BG724280.1	GI:14003454		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	1	(bases 1 to 625)		
AUTHORS	NIH-MGC	http://mgc.nci.nih.gov/.		
TITLE	National Institutes of Health,			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D.			
	Email: cgapbs-r@mail.nih.gov			
	Tissue Procurement: Miklos Palkovits, M.D., Ph.D.			
	CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki			
	Toshiyuki and Piero Carninci (RIKEN)			
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
	DNA Sequencing by: Incyte Genomics, Inc.			
	Clone distribution: MGC clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LLNL at:			
	http://image.llnl.gov			
	Plate: LLAM10750	row: e	column: 02	
	High quality sequence stop: 623.			
FEATURES	Location/Qualifiers			
source	1. .625			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:4829929"			
	/clone_lib="NIH_MGC_97"			
	/lab_host="DH10B"			
	/note="Organ: testis; Vector: pBluescriptR (modified			
	pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag			
	); Oligo-dT primed using primer 5'-TTTTTTT			
	size-selected for average insert size 2.2 kb and			
	normalized to ROT 5. This is a primary library enriched			
	for full-length clones and constructed using the			
	Cap-trapper method (Carninci, in preparation). Library			
	constructed by M. Brownstein (NIMH/NHGRI, National			
	Institutes of Health). Note: this is a NIH_MGC Library."			
BASE COUNT	172 a	134 c	156 g	163 t
ORIGIN				
Query Match	14.0%;	Score 498.4;	DB 12;	Length 625;
Best Local Similarity	99.8%;	Pred. NO. 3.6e-90;		
Matches 497;	Conservative	0;	Mismatches	1;
			Indels	0;
			Gaps	0;
QY	1393	GGTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGTTCTGACATGATTAC	1452	
Db	115	GGTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGTTCTGACATGATTAC	174	
QY	1453	CTGATCTTAAACTTGCCATATCCAGGATTTAAACCACCAACCATCTATGACTGGAGCCACG	1512	
Db	175	CTGATCTTAAACTTGCCATATCCAGGATTTAAACCACCAACCATCTATGACTGGAGCCACG	234	

QY	1513	GGTTCTGAAC	TAGGAAGAAATAACATTTTGCTCTTTGAGACCCCTCTGTTCACTGACTGTGTT	1572
Db	235	GGTTCTGAAC	TAGGAAGAAATAACATTTTGCTCTTTGAGACCCCTCTGTTCACTGACTGTGTT	294
QY	1573	TTGTACTT	CATGGTGGATATTAAATAGAAAAAGTACAAAATGGTAGAATCGTGGGTGGA	1632
Db	295	TCGTACTT	CATGGTGGATATTAAATAGAAAAAGTACAAAATGGTAGAATCGTGGGTGGA	354
QY	1633	ACCAAAGAAAAA	CAAGCTTACACCCCATATCATCTTCAAGAATGCAACTTTTACATTTACA	1692
Db	355	ACCAAAGAAAAA	CAAGCTTACACCCCATATCATCTTCAAGAATGCAACTTTTACATTTACA	414
QY	1693	TGGGCATT	CCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTCATCAATGACATGGTG	1752
Db	415	TGGGCATT	CCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTCATCAATGACATGGTG	474
QY	1753	AAGATTTAT	TCTATCACAGCCACTAATGCAGTTGATGGGGTGGCGTCCCTCATGCCGTGCC	1812
Db	475	AAGATTTAT	TCTATCACAGCCACTAATGCAGTTGATGGGGTGGCGTCCCTCATGCCGTGCC	534
QY	1813	TGTGCCCT	CGGTTCTGAACAGTCGGGTTTCATCGTGTGCCCTGCCCTCCAGGCCACTAC	1872
Db	535	TGTGCCCT	CGGTTCTGAACAGTCGGGTTTCATCGTGTGCCCTGCCCTCCAGGCCACTAC	594
QY	1873	ATTGAGAAAGAA	CAAC	1890
Db	595	ATTGAGAAAGAA	CAAC	612
RESULT 15				
BI731892				
LOCUS	BI731892	801 bp	mRNA	linear EST 20-SEP-2001
DEFINITION	603353640F1 NIH_MGC_94	Mus musculus	CDNA clone	IMAGE:5361249 5', mRNA sequence.
ACCESSION	BI731892			
VERSION	BI731892.1	GI:15708905		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	1	(bases 1 to 801)		
AUTHORS	NIH-MGC	http://mgc.nci.nih.gov/.		
TITLE	National Institutes of Health,			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D.			
	Email: cgapbs-r@mail.nih.gov			
	Tissue Procurement: The Cepko Laboratory			
	CDNA Library Preparation: Life Technologies, Inc.			
	DNA Sequencing by: Incyte Genomics, Inc.			
	Clone distribution: MGC clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LLNL at:			
	http://image.llnl.gov			
	Plate: LLAM11918	row: o	column: 10	
	High quality sequence stop: 797.			
FEATURES	Location/Qualifiers			
source	1. .801			
	/organism="Mus musculus"			
	/db_xref="taxon:10090"			
	/clone="IMAGE:5361249"			
	/clone_lib="NIH_MGC_94"			
	/tissue_type="retina"			
	/lab_host="DH10B (phage-resistant)"			
	/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;			
	Site_2: SalI; Cloned unidirectionally; oligo-dT primed.			
	Average insert size 3.3 kb. Library enriched for			
	full-length clones and constructed by Life technologies.			
	Note: this is a NIH_MGC Library."			
BASE COUNT	241 a	161 c	197 g	202 t
ORIGIN				
Query Match	13.9%;	Score 494.2;	DB 13;	Length 801;

Best Local Similarity 85.7%; Pred. No. 2.4e-89;									
Matches 630;		Conservative 0;		Mismatches 98;		Indels 7;		Gaps 7;	
QY	2375	TGAAAAATATTAAATATAAAAGAAGATATGTTCCCGAGTTC	CAACAAGCCAAATACCAGATG	2434					
Db	10	TGCAAAACGTTAATATCAAGGAGGATATGTGCCCAGGT-CACCGAGCCAA	GTACCAGATG	68					
QY	2435	TGCATTTCTTTTATAAGTCTTCTACAGCAACAACATCTTGTATTAA	TGGCCGATCAACTG	2494					
Db	69	TTCATTTCTTTTATAAGTCTTCCACAGCCACCACATC-TGTATTAA	TGGACGATCGACTG	127					
QY	2495	CTGTGAAAATGAGGTGTAATCCTTACTAAATCTGGAGCAGGAGTGAT	TTTCAGTCCCCAGCA	2554					
Db	128	CTGTGAAGATGAGGTGTAATCCCATGAGACCTGGTGAGGTGTGATAT	CAGTCCCCAGCA	187					
QY	2555	AGTGCCCGACGAGGTACCT-GTGATGGGTGTACGTTCTATTTCCTGT	GGGAGAGTGCTGAA	2613					
Db	188	AGTGTCAGCTGGCACCTGGTGATGGTTGTACCTTCTACTTTCTAT	GGGAGAGTGCAGAA	247					
QY	2614	GCTTGCCCTCTGTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCC	TGCAAGAGAGGA	2673					
Db	248	GCTTGCCCTCTGTGCACAGAACATGACTTCCATGAGATCGAGGGAGC	CTGCAAGAGAGGG	307					
QY	2674	TTTCAGGAAACCTTGTATGTGTGGAATGAACCTAAATGGTGCAATT	AAAGGAATTC	2733					
Db	308	CTTCAGGAAATATTATATGTATGGAATGAACCTTAAGTGGTGCA	TTAAAGGAATTC	367					
QY	2734	CCTGAGAAAAAGTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGA	AGGTGGGAGCCGGT	2793					
Db	368	CCCGAAAGAAGTTGTCAACCTGCGAAACTGTTGAC-TTTGGCTGA	AAAGTGGGAGCTGGT	426					
QY	2794	GTGGGAGCTTTTACTGCGGTTTGTCTGGTGGCTCTGACCTGCTAC	TTCTGGAAAAAGAAT	2853					
Db	427	GTGGGCGGTTTCACAGCGG-TTTGTTGGTGGCTTTAACATGCTAC	TTTGGAAAAAGAAT	485					
QY	2854	CAAAAACCTGGAATACAAATATCCAAAGTTAGTAATGACGACTAAC	CTCAAAAGAGTGTGAA	2913					
Db	486	CAAAAGCTGGAATACAAATATTTAAATTAGTAATGACGACTAACT	CAAAAGAGTGTGAA	545					
QY	2914	CTCCCGGCTGCAGACAGTGTGTCTATCATGGAAGGAGAAAGATA	ATGAAGAGGAAGTTGTA	2973					
Db	546	CTCCAGCTGCAGACAGCTGTGTCTATCATGGAAGGAGAAAGATA	ATGAAGAGGATGTGGTA	605					
QY	2974	TATTCCAATAAACAGTCACTACTAGGAAAACTCAAATCTTTGGCA	ACCACAGGAAAAAGAA	3033					
Db	606	TATTCCAACAAACAGTCACTACTAGGAAAACTTAAGTCTTGGCC	ACAAAGSAAAAAAGAT	665					
QY	3034	GACCATTTTGAATCTGTCAACTGAAAACCTCAACATCCCCCAA	ATATATGACAGAC-AG	3092					
Db	666	GACCA-CTTGAGTCTGTTCAGCTGAAATCCCTCAAGATGCCCA	GTATATGAAGAGACAAA	724					
QY	3093	TGCTGTAGCCTTGAG	3107						
Db	725	TGCTGTAGCATTCAG	739						

Search completed: May 12, 2003, 07:48:31  
Job time : 4363.66 secs

GenCore version 5.1.5  
 Copyright (c) 1993 - 2003  Compugen Ltd.  
  
 OM nucleic - nucleic search, using sw model  
  
 Run on:           May 12, 2003, 00:12:42 ; Search time 4131.84 Seconds  
                   (without alignments)  
                   13287.722 Million cell updates/sec  
  
 Title:            US-10-073-333A-1  
 Perfect score:    3390  
 Sequence:         1 atgctgttcgcgcgccgggg.....ggtaaaaaaaaaaaaaaaaa 3390  
  
 Scoring table:    IDENTITY\_NUC  
                  Gapop 10.0 , Gapext 1.0  
  
 Searched:         16154066 seqs, 8097743376 residues  
  
 Total number of hits satisfying chosen parameters:    32308132  
  
 Minimum DB seq length: 0  
 Maximum DB seq length: 20000000000  
  
 Post-processing: Minimum Match 0%  
                  Maximum Match 100%  
                  Listing first 45 summaries  
  
 Database :        EST:\*  
                  1: em\_estba:\*  
                  2: em\_esthum:\*  
                  3: em\_estin:\*  
                  4: em\_estmu:\*  
                  5: em\_estov:\*  
                  6: em\_estpl:\*  
                  7: em\_estro:\*  
                  8: em\_htc:\*  
                  9: gb\_est1:\*  
                  10: gb\_est2:\*  
                  11: gb\_htc:\*  
                  12: gb\_est3:\*  
                  13: gb\_est4:\*  
                  14: gb\_est5:\*  
                  15: em\_estfun:\*  
                  16: em\_estom:\*  
                  17: gb\_gss:\*  
                  18: em\_gss\_hum:\*  
                  19: em\_gss\_inv:\*  
                  20: em\_gss\_pln:\*  
                  21: em\_gss\_vit:\*  
                  22: em\_gss\_fun:\*  
                  23: em\_gss\_mam:\*  
                  24: em\_gss\_mus:\*  
                  25: em\_gss\_other:\*  
                  26: em\_gss\_pro:\*  
                  27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	660.4	19.5	697 14 BM723980
2	654.4	19.3	656 14 BM783507
3	624.8	18.4	930 12 BF035563
4	610.4	18.0	646 10 AW954806
5	553.2	16.3	762 13 BI227132
6	534.6	15.8	705 14 BQ182610

SUMMARIES			
Result No.	Score	Query Match	Description
1	660.4	19.5	697 14 BM723980
2	654.4	19.3	656 14 BM783507
3	624.8	18.4	930 12 BF035563
4	610.4	18.0	646 10 AW954806
5	553.2	16.3	762 13 BI227132
6	534.6	15.8	705 14 BQ182610

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	660.4	19.5	697 14 BM723980
2	654.4	19.3	656 14 BM783507
3	624.8	18.4	930 12 BF035563
4	610.4	18.0	646 10 AW954806
5	553.2	16.3	762 13 BI227132
6	534.6	15.8	705 14 BQ182610

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	660.4	19.5	697 14 BM723980
2	654.4	19.3	656 14 BM783507
3	624.8	18.4	930 12 BF035563
4	610.4	18.0	646 10 AW954806
5	553.2	16.3	762 13 BI227132
6	534.6	15.8	705 14 BQ182610

SUMMARIES			
Result No.	Score	Query Match	Description
7	527	15.5	527 14 BM752466
8	513.4	15.1	667 10 BB384282
9	512.8	15.1	634 14 BQ551364
10	507.4	15.0	515 14 BM771726
11	498.4	14.7	625 12 BG724280
12	493	14.5	541 10 BE237235
13	492.8	14.5	713 14 BQ772508
14	482.2	14.2	557 14 W60843
15	482.2	14.2	576 14 W60844
16	481.4	14.2	484 9 AA010992
17	478	14.1	478 10 AW665873
18	477	14.1	477 9 AI885785
19	472.6	13.9	585 14 BM974911
20	463	13.7	896 12 BG116864
21	449.4	13.3	508 13 BM538241
22	446	13.2	458 12 BF740039
23	435.6	12.8	1109 14 BQ896007
24	424.8	12.5	600 10 BB617824
25	421.8	12.4	656 10 BB626387
26	419.2	12.4	653 10 BB650824
27	414.4	12.2	507 9 AI797353
28	404	11.9	420 13 BI480715
29	397	11.7	445 9 AA010869
30	394.4	11.6	641 10 BB629056
31	391.8	11.6	667 13 BM491171
32	388.6	11.5	640 10 BB622295
33	381.8	11.3	859 14 BQ730700
34	374.6	11.1	449 14 HI8553
35	374	11.0	374 9 AA644099
36	372	11.0	1773 11 AK017241
37	370.8	10.9	446 9 AI317059
38	369.4	10.9	395 10 AW444960
39	344.4	10.2	553 10 BB644517
40	332.6	9.8	631 13 BJ069789
41	331.6	9.8	387 10 BE081050
42	328.4	9.7	433 9 AI982900
43	325.2	9.6	801 13 BI731892
44	322.4	9.5	695 10 BB651047
45	317.2	9.4	689 9 AL644075

ALIGNMENTS

RESULT 1	BM723980	BM723980	697 bp	mRNA	linear	EST 01-MAR-2002
LOCUS	UI-E-E01-aiy-a-05-0-UI.r1	UI-E-E01 Homo sapiens	cdNA clone			
DEFINITION	UI-E-E01-aiy-a-05-0-UI 5', mRNA sequence.					
ACCESSION	BM723980					
VERSION	BM723980.1	GI:19045304				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1 (bases 1 to 697)					
TITLE	Bonaldo,M.F., Lennon,G. and Soares,M.B.					
JOURNAL	Normalization and subtraction: two approaches to facilitate gene discovery					
MEDLINE	Genome Res. 6 (9), 791-806 (1996)					
COMMENT	97044477					
	Contact: Soares, MB					
	Program for Rat Gene Discovery and Mapping					
	University of Iowa					
	451 Eckstein Medical Research Building Iowa City, IA 52242, USA					
	Tel: 319 335 8250					
	Fax: 319 335 9565					
	Email: msoares@blue.weeg.uiowa.edu					
	Tissue Procurement: Dr. Gregg Hageman					
	cdNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa					
	cdNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa					
	DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa					

SUMMARIES			
Result No.	Score	Query Match	Description
1	660.4	19.5	697 14 BM723980
2	654.4	19.3	656 14 BM783507
3	624.8	18.4	930 12 BF035563
4	610.4	18.0	646 10 AW954806
5	553.2	16.3	762 13 BI227132
6	534.6	15.8	705 14 BQ182610



Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. .697  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-E-EO1-aiy-a-05-0-UI"  
/clone\_lib="UI-E-EO1"  
/tissue\_type="fetal eye"  
/dev\_stage="fetal"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-E-EO1 is a normalized cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGCGTATACC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."  
BASE COUNT 222 a 150 c 138 g 186 t 1 others  
ORIGIN  
Query Match 19.5%; Score 660.4; DB 14; Length 697;  
Best Local Similarity 99.4%; Pred. No. 2.4e-122;  
Matches 694; Conservative 0; Mismatches 1; Indels 3; Gaps 3;  
QY 1835 CGGGTTCATCGTGTGCCCTGCCCTCCAGGCCACTACAT-TGAGAAAGAAACCAACCAG 1893  
|||||  
Db 1 CGGGTTCATCGTGTGCCCTGCCCTCCAGGCCACTACATNTGAGAAAGAAACCAACCAG 60  
QY 1894 TGCAAGGAATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGAGGCT 1953  
|||||  
Db 61 TGCAAGGAATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGAGGCT 120  
QY 1954 TGTATTCCATCGGGGCCCTGGGAGTAAACAACATCAGGACCATTTCGGTTTGCTATAGTGAC 2013  
|||||  
Db 121 TGTATTCCATCGGGGCCCTGGGAGTAAACAACATCAGGACCATTTCGGTTTGCTATAGTGAC 180  
QY 2014 TGCTTTTCTACCATGAAAAAGAAATCAGATTTTGCACATGACTTTAGCAACCTCAGC 2073  
|||||  
Db 181 TGCTTTTCTACCATGAAAAAGAAATCAGAGTTTGCACTATGACTTTAGCAACCTCAGC 240  
QY 2074 AGTGTGGGCTCATTAATGAATGGCCCGAGCTTCACCTCCAAAGGAACAAATACTTCCAT 2133  
|||||  
Db 241 AGTGTGGGCTCATTAATGAATGGCCCGAGCTTCACCTCCAAAGGAACAAATACTTCCAT 300  
QY 2134 TTCTTCAATATCAGTTTATCTGGGCATGAGGGGAAGAGATGSGTCTCTGTACCAACAAT 2193  
|||||  
Db 301 TTCTTCAATATCAGTTTATCTGGGCATGAGGGGAAGAGATGSGTCTCTGTACCAACAAT 360  
QY 2194 ATAACAGACTTTACAGTAAAGAAATAGTGGCAGGGTCAGATGATTACACAAA-TTTGGT 2252  
|||||  
Db 361 ATAACAGACTTTACAGTAAAGAAATAGTGGCAGGGTCAGATGATTACACAAAATTTGGT 420  
QY 2253 AGGGGCATTTGTATGCCAGTCAACAATATTCTTCTGAAAGTAAGGGTTTCCGAGCAGC 2312  
|||||  
Db 421 AGGGGCATTTGTATGCCAGTCAACAATATTCTTCTGAAAGTAAGGGTTTCCGAGCAGC 480  
QY 2313 CTTATCATCACAATCCATCATTTCTGCGAGATACATTTCATAGGAGTCACAGTTGAAACCCAC 2372  
|||||  
Db 481 CTTATCATCACAATCCATCATTTCTGCGAGATACATTTCATAGGAGTCACAGTTGAAACCCAC 540  
QY 2373 ATTGAAAAATATTAAATAAAAGAGATATGTTCCCAGTTTCCCAACAGCCAAATACCAGA 2432  
|||||

Db 541 ATTGAAAAATATTAATAATAAAGAAGATATGTTCCCAAGTTCACCAAGCCAAATACCAGA 600  
QY 2433 TGTGCATTTCTTTTATAAGTCTTCTACAGCAACAACATCTTGTATTAAATGGCCGATCAAC 2492  
|||||  
Db 601 TGTGCATTTCTTTTATAAGTCTTCTACAGCAA-AACATCTTGTATTAAATGGCCGATCAAC 659  
QY 2493 TGCCTGTAANAATGAGGTGTAATCCTACTACTAAATCTGGAG 2530  
|||||  
Db 660 TGCCTGTAANAATGAGGTGTAATCCTACTACTAAATCTGGAG 697  
RESULT 2  
BM783507 656 bp mRNA linear EST 05-MAR-2002  
LOCUS K-EST0061445 S7SNU719 Homo sapiens cDNA clone S7SNU719-36-H08 5',  
DEFINITION mRNA sequence.  
ACCESSION BM783507  
VERSION BM783507.1 GI:19131739  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Plrimates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 656)  
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished (2002)  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 36 row: H column: 08  
High quality sequence stop: 656.  
Location/Qualifiers  
1. .656  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="S7SNU719-36-H08"  
/clone\_lib="S7SNU719"  
/sex="M"  
/tissue\_type="Stomach"  
/cell\_type="Epithelial"  
/cell\_line="SNU-719"  
/lab\_host="Top10F"  
/note="Organ: Stomach; Vector: pCNS; Site\_1: EcoRI;  
Site\_2: NotI; The poly (A)+ RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then decapped  
with tabacco acid pyrophosphatase (TAP). The decapped  
intact mRNA was ligated with DNA-RNA linker including EcoR  
I site by treatment of T4 RNA ligase and the first strand  
cDNA was synthesized from oligo dT-selected mRNA by  
priming with dT-tailed vector. The dT-tailed vector was  
adjusted to have about 60nt. The cDNA vector was  
circularized with E. coli DNA ligase after digestion of  
EcoRI which site is also included in vector. An RNA strand  
converted to a DNA strand by Okayama-Berg method. The  
obtained cDNA vectors were used for transformation of  
competent cells E. coli Top10F' by electroporation method.  
The cDNA libraries constructed by this method are  
full-length enriched cDNA library."  
BASE COUNT 191 a 147 c 145 g 173 t  
ORIGIN  
Query Match 19.3%; Score 654.4; DB 14; Length 656;  
Best Local Similarity 99.8%; Pred. No. 3.9e-121;  
Matches 655; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1710 TAATCAGGGTCAAGATAATAGACGGTTCATCAATGACATGGTGAAGATTATTCTATC 1769

```
Db 1 TAATCAGGGTCAAGATAATAGACGGTTTCATCAATGACATGSGAAGATTATTCTATCAC 60
QY 1770 AGCCACTAATGCAGTTGATGGGGTGGCGTCTCTCATGCCGTGCTGCCCTCGGTTCTGA 1829
Db 61 AGCCACTAATGCAGTTGATGGGGTGGCGTCTCTCATGCCGTGCTGCCCTCGGTTCTGA 120
QY 1830 ACAGTCGGGTTTCATCGTGTGTCCTCCCTGCCCTCCAGGCCACTACATTGAGAAAGAAACCAA 1889
Db 121 ACAGTCGGGTTTCATCGTGTGTCCTCCCTGCCCTCCAGGCCACTACATTGAGAAAGAAACCAA 180
QY 1890 CCAGTGAAGGAATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGA 1949
Db 181 CCAGTGAAGGAATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGA 240
QY 1950 GGCTTGTATTCCATCGCGGCCCTGGGAGTAAAAACAATCAGGACCATTGGTTTGCTATAG 2009
Db 241 GGCTTGTATTCCATCGCGGCCCTGGGAGTAAAAACAATCAGGACCATTGGTTTGCTATAG 300
QY 2010 TGACTGCTTTTCTTACCATTGAAAAAGAAAAATCAGATTTTGCACATGACTTTAGCAACCT 2069
Db 301 TGACTGCTTTTCTTACCATTGAAAAAGAAAAATCAGAGTTTGCACATGACTTTAGCAACCT 360
QY 2070 CAGCAGTGTGGGCTCATTAATGAATGCCCCCAGCTTCACCTCCAAAGGAACAAAATACTT 2129
Db 361 CAGCAGTGTGGGCTCATTAATGAATGCCCCCAGCTTCACCTCCAAAGGAACAAAATACTT 420
QY 2130 CCATTCTTCAATATCAGTTTATGTGGGCATGAGGGGAAGAGATGGCTCTCTGTACCAA 2189
Db 421 CCATTCTTCAATATCAGTTTATGTGGGCATGAGGGGAAGAGATGGCTCTCTGTACCAA 480
QY 2190 CAATATAACAGACTTTACAGTAAAAAGAAATAGTGGCAGGGTCAGATGATTACACAAATTT 2249
Db 481 CAATATAACAGACTTTACAGTAAAAAGAAATAGTGGCAGGGTCAGATGATTACACAAATTT 540
QY 2250 GGTAGGGGCATTTGTATGCCAGTCAACAATATTATTCCTTCTGAAAGTAAGGGTTTCCGAGC 2309
Db 541 GGTAGGGGCATTTGTATGCCAGTCAACAATATTATTCCTTCTGAAAGTAAGGGTTTCCGAGC 600
QY 2310 AGCCTTATCATCACAAATCCATCATCTTCTGGCAGATACATTTCATAGGAGTCACAGTTG 2365
Db 601 AGCCTTATCATCACAAATCCATCATCTTCTGGCAGATACATTTCATAGGAGTCACAGTTG 656
```

```
RESULT 3
BF035563
LOCUS
DEFINITION
  BF035563
  601457452F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3861156 5',
  mRNA sequence.
ACCESSION
  BF035563
VERSION
  BF035563.1 GI:10743303
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 930)
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: DCTD/DTP
  CDNA Library Preparation: Life Technologies, Inc.
  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLAM9597 row: g column: 13
  High quality sequence stop: 647.
FEATURES
  source
    Location/Qualifiers
      1. .930
```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3861156"
/clone_lib="NIH_MGC_66"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT      275 a      222 c      211 g      222 t
ORIGIN
Query Match      18.4%; Score 624.8; DB 12; Length 930;
Best Local Similarity 91.4%; Pred. No. 3.le-115;
Matches 684; Conservative 0; Mismatches 62; Indels 2; Gaps 2;
QY 1661 TCATCTTCAAGAATGCAACTTTTACATTTACATGGGCATTCAGAGAACTAATCAGGGTC 1720
Db 1 TCATCTTCAAGAATGCAACTTTTACATTTACATGGGCATTCAGAGAACTAATCAGGGTC 60
QY 1721 AAGATAATAGACGGTTCATCAATGACATGGTGAAGATTTATTCTATCACAGCCACTAATG 1780
Db 61 AAGATAATAGACGGTTCATCAATGACATGGTGAAGATTTATTCTATCACAGCCACTAATG 120
QY 1781 CAGTTGATGGGGTGGCGTCCCTCATGCCGTGCCTGTGCCCTCGGTTCTGAACAGTCGGGT 1840
Db 121 CAGTTGATGGGGTGGCGTCCCTCATGCCGTGCCTGTGCCCTCGGTTCTGAACAGTCGGGT 180
QY 1841 CATCGTGTCCCCCTGCCCTCCAGGCCACTACATTGAGAAAGAAACCAACCAAGTCAAGG 1900
Db 181 CATCGTGTCCCCCTGCCCTCCAGGCCACTACATTGAGAAAGAAACCAACCAAGTCAAGG 240
QY 1901 AATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGAGGCTTGATTTC 1960
Db 241 AATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGAGGCTTGATTTC 300
QY 1961 CATGGGGCCTGGGAGTAAAAACAATCAGGACCATTCCGTTTGCTATAGTACTGCTTTT 2020
Db 301 CATGGGGCCTGGGAGTAAAAACAATCAGGACCATTCCGTTTGCTATAGTACTGCTTTT 360
QY 2021 TCTACCATGAAAAAGAAATCAGATTTTGCACATGACTTTAGCAACCTCAGCAGTGTGG 2080
Db 361 TCTACCATGAAAAAGAAATCAGAGTTTGCACATGACTTTTAGCAACCTCAGCAGTGTGG 420
QY 2081 GCTCATTAAATGAATGGCCCCAGCTTCACCTCCAAAGGAACAAAATACTTCCATTCTTCA 2140
Db 421 GCTCATTAAATGAATGGCCCCAGCTTCACCTCCAAAGGAACAAAATACTTCCATTCTTCA 480
QY 2141 ATATCAGTTTATGTGGGCATGAGGGGAAGAAAGATGGCTCTCTGTACCAACAATATAACAG 2200
Db 481 ATATCAGTTTATGTGGGCATGAGGGGAAGAAAGATGGCTCTCTGTACCAACAATATAACAG 540
QY 2201 ACTTTACAGTAAAAAGAAATAGTGGCAGGGTCAGATGATTACACAAATTTGGTAGGGGCAT 2260
Db 541 ACTTTACAGTACAAGACATAGTGGCA-GGTCAGATGATTACACAAATTTGGTAGGGGCATT 599
QY 2261 TTGTATGCCAGTCAACAATTTATTCCTTCTGAAAGTAAGGGTTTCCGAGCAGCCTTATCAT 2320
Db 600 GTATGCCAGTCAACAATTTATTCCTTCTGAAAGTAAGGGTTTCCGAGCAGACTTGTATCAT 659
QY 2321 CACAATCCA-TCATTCTGGCAGATACATTCATAGGAGTCACAGTTGAAACCCACATTTGAAA 2379
Db 660 CACCATACATTCATTCTGGCGGATACATCTCTAGGAGTCACAGGTTGAAACCCACATTTGGA 719
QY 2380 AATATTAAATATAAAGAAAGATATGTTC 2407
Db 720 AACTTTCATTTCACAGAAGATATGTCCC 747
RESULT 4
AW954806
LOCUS
  AW954806
  646 bp      mRNA      linear      EST 01-JUN-2000
```

DEFINITION EST366876 MAGE resequences, MAGC Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW954806  
VERSION AW954806.1 GI:8144489  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 646)  
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and Quackenbush,J.  
TITLE Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray  
JOURNAL Unpublished (2000)  
COMMENT Contact: John Quackenbush  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johnq@tigr.org  
Plate: 69  
Seq primer: Reverse.  
Location/Qualifiers  
1..646  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MAGE resequences, MAGC"  
/note="Vector: pBluescriptSKm"  
BASE COUNT 189 a 143 c 142 g 171 t 1 others  
ORIGIN  
Query Match 18.0%; Score 610.4; DB 10; Length 646;  
Best Local Similarity 98.8%; Pred. No. 2.7e-112;  
Matches 636; Conservative 0; Mismatches 6; Indels 2; Gaps 2;  
QY 1440 TGACAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCACCAACATCTAT 1499  
Db 1 TGACAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCACCAACATCTAT 60  
QY 1500 GACTGGAGCCACGGGTTCTGAACTAGGAAGAATAACATTTGTCTTTGAGACCCCTCTGTC 1559  
Db 61 GACTGGAGCCACGGGTTCTGAACTAGGAAGAATAACATTTGTCTTTGAGACCCCTCTGTC 120  
QY 1560 AGCTGACTGTGTTTGTACTTCATGGTGGATATTAAATAGAAAAAGTACAAATGGGTAGA 1619  
Db 121 AGCTGACTGTGTTTGTACTTCATGGTGGATATTAAATAGAAAAAGTACAAATGGGTAGA 180  
QY 1620 ATCGTGGGTGGAACCAAGAAACAAGCTTACACCCCATATCATCTTCAAGAATGCAAC 1679  
Db 181 ATCGTGGGTGGAACCAAGAAACAAGCTTACACCCCATATCATCTTCAAGAATGCAAC 240  
QY 1680 TTTTACATTTACATGGGCATTCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTTAT 1739  
Db 241 TTTTACATTTACATGGGCATTCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTTAT 300  
QY 1740 CAATGACATGGTGAAGATTTATTCTATCACAGCCACTAATGCAGTTGATGGGGTGGCGTC 1799  
Db 301 CAATGACATGGTGAAGATTTATTCTATCACAGCCACTAATGCAGTTGATGGGGTGGCGTC 360  
QY 1800 CTCATGCGGTGCCTGTGCCCTCGGTTCTGAACAGTCGGGTTTCATCGTGTGTCCCTGCC 1859  
Db 361 CTCATGCGGTGCCTGTGCCCTCGGTTCTGAACAGTCGGGTTTCATCGTGTGTCCCTGCC 420  
QY 1860 TCCAGGCCACTACATTTAGAGAAAGAAACCAACCAAGTGAAGGAATGTCCACCTGACACCTA 1919  
Db 421 TCCAGGCCACTACATTTAGAGAAAGAAACCAACCAAGTGAAGGAATGTCCACCTGACACCTA 480  
QY 1920 CCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGATTTCCATGCGGGCCTGGGAGTAA 1979  
Db 481 CCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGATTTCCATGCGGGCCTGGGAGTAA 540  
QY 1980 AAACAATCAGGACCATTCGGGTTTGCTATAGTGACTGCTTTTCTTACCATG-AAAAAGAA 2038

Db 541 AAACAATCAGGACCCCTTCGGTTTGTATAGTGACTGCTTTTCTACCATGAAAAAGAAA 600  
QY 2039 ATCAGATTTTGCACATATGAC-TTTAGCAACCTCAGCAGTGTGGG 2081  
Db 601 ATCAGAGTTTGCACATATGACNTTTGGCAACCTTAACAGGGTGGG 644  
RESULT 5  
BI227132 762 bp mRNA linear EST 11-JUL-2001  
LOCUS 602952315F1 NIH\_MGC\_8 Homo sapiens cDNA clone IMAGE:5096643 5',  
DEFINITION mRNA sequence.  
ACCESSION BI227132  
VERSION BI227132.1 GI:14680576  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 762)  
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM1868 row: n column: 04  
High quality sequence stop: 659.  
Location/Qualifiers  
1..762  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5096643"  
/clone\_lib="NIH\_MGC\_8"  
/tissue\_type="Burkitt lymphoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lymph; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 225 a 176 c 178 g 183 t  
ORIGIN  
Query Match 16.3%; Score 553.2; DB 13; Length 762;  
Best Local Similarity 96.2%; Pred. No. 7.7e-101;  
Matches 612; Conservative 0; Mismatches 13; Indels 11; Gaps 4;  
QY 890 CATTCCAGCAACAAACCAGGTTTCATTCAACTGCCAGGTGTGCCAGAAACACCTATTCTG 949  
Db 2 CATTCCAGCAACAAACCAGGTTTCATTCAACTGCCAGGTGTGCCAGAAACACCTATTCTG 61  
QY 950 AGAAGGAGCCCAAGAATGTATAAGTGTAAAGACGACTCTCAATTTTC-----AGGAT 1003  
Db 62 AGAAGGAGCCCAAGAATGTATAAGTGTAAAGACGACTCTCAATTTTCAGAGGAAGGAT 121  
QY 1004 CCAGTGAGTGACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATATCTC 1063  
Db 122 CCAGTGAGTGACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATATCTC 181  
QY 1064 CATGTGATGAAGAGGAAAGACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCC 1123  
Db 182 CATGTGATGAAGAGGAAAGACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCC 241





AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001

JOURNAL Unpublished (2002)

COMMENT Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 17 row: H column: 12

High quality sequence stop: 527.

Location/Qualifiers

source 1. .527

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="S9SNU601-17-H12"

/clone\_lib="S9SNU601"

/sex="M"

/tissue\_type="Ascites"

/cell\_type="Epithelial"

/cell\_line="SNU-601"

/lab\_host="Top10F"

/note="Organ: Stomach; Vector: pME18-FL3; Site\_1: XhoI; Site\_2: XhoI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including SfiI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized with Superscript II using SfiI oligo-dT primer. After first strand synthesis, RNA was degraded by NaOH treatment and cDNA was amplified by PCR reaction. The PCR products were digested with SfiI and cloned into DraIII- digested pME18S-FL3 vector. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

BASE COUNT 156 a 120 c 116 g 135 t

ORIGIN

Query Match 15.5%; Score 527; DB 14; Length 527;

Best Local Similarity 100.0%; Pred. No. 1.6e-95;

Matches 527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1461 AAACCTGCATATCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGGTTCTGA 1520

|||||

Db 1 AAACCTGCATATCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGGTTCTGA 60

QY 1521 ACTAGGAAGAATAACATTTGTCTTTGAGACCCCTCTGTTTCAGCTGACTGTGTTTGTACTT 1580

|||||

Db 61 ACTAGGAAGAATAACATTTGTCTTTTGAGACCCCTCTGTTTCAGCTGACTGTGTTTGTACTT 120

QY 1581 CATGGTGGATATTAAAGAAAAAGTACAAATGTGGTAGAATCGTGGGGTGGAAACCAAGA 1640

|||||

Db 121 CATGGTGGATATTAAAGAAAAAGTACAAATGTGGTAGAATCGTGGGGTGGAAACCAAGA 180

QY 1641 AAAACAAGTTACACCCCATATCATCTTTCAAGAATGCAACTTTTACATTTACATGGGCATT 1700

|||||

Db 181 AAAACAAGTTACACCCCATATCATCTTTCAAGAATGCAACTTTTACATTTACATGGGCATT 240

QY 1701 CCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTTCATCAATGACATGGTGAAGATTTA 1760

|||||

Db 241 CCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTTCATCAATGACATGGTGAAGATTTA 300

QY 1761 TTCTATCAGACCCACTAATGCAGTTTGATGGGGTGGCGTCCCTCATGCCGTGCTGCCCCCT 1820

|||||

Db 301 TTCTATCAGACCCACTAATGCAGTTTGATGGGGTGGCGTCCCTCATGCCGTGCTGCCCCCT 360

QY 1821 CGGTTCTGACAGTCGGGTTTCATCTGTGTGCCCTCCAGCCCACTACATTGAGAA 1880

|||||

Db 361 CGGTTCTGAACAGTCGGGTTTCATCGTGTGTCCCTGCCCTCCAGGCCACTACATTGAGAA 420

QY 1881 AGAAACCAACCAGTGTCAAGGAATGTCCACCTGACACCTACCTACATCATCATCAGGTCTTA 1940

|||||

Db 421 AGAAACCAACCAGTGTCAAGGAATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTTA 480

QY 1941 TGGCAAAAGAGGCTTGTATTTCATGCGGGCCTGGGAGTAAACAAATC 1987

|||||

Db 481 TGGCAAAAGAGGCTTGTATTTCATGCGGGCCTGGGAGTAAACAAATC 527

RESULT 8

BB384282

LOCUS

DEFINITION BB384282 RIKEN full-length enriched, 0 day neonate cerebellum Mus musculus CDNA clone C230033C19 3', mRNA sequence.

ACCESSION BB384282

VERSION BB384282.2 GI:16408254

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 667)

AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

JOURNAL Unpublished (2001)

COMMENT On Jul 13, 2000 this sequence version replaced gi:9104048. Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL:http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Location/Qualifiers

FEATURES





QY 1154 CCCCTTCTGGAGAGAAGGATTGTCCGCCCTTGCAACCCCTGGATTTTATAACAATGGAT 1213  
Db 241 CCCTTCTGGAGAGAAGGATTGTCTCTTGGCAACCCAGGATTCTATAACAATGGAT 300  
QY 1214 CATCTTCTGCCATCCCTGTCTCTCTGGAAACATTTTCAGATGGAACCAAGAATGTAGAC 1273  
Db 301 CGTCTTCTGCCATCCCTGCCCTCCGGGGACGTTTTCAGATGGAACCAAGAATGCAAGT 360  
QY 1274 CATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAATGGTGAATGTCCTTC 1333  
Db 361 CATGTCCGGCCGGAACAGAGCCAGCACCTGGGTTTGAATATAAATGGTGAATGTCCTCC 420  
QY 1334 CTGGCAACATGAAAACTTCTCTGCTTCAATGTTGGGAATTCAAAGTCCGATGGAATGAATG 1393  
Db 421 CTGCCAACATGAAAACTTCTCTGCTTCAATGTTGGGAATTCCAAGTCCGATGGAATGAACG 480  
QY 1394 GTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGACAATGATTACC 1453  
Db 481 GTTGGGAGGTGGCTGGAGATCATATCCGGAGTGGAGCTGGAGGCTCTGACAATGACTATC 540  
QY 1454 TGATCTTAAACTTGCATATCCCAGGATTTAAACCAACCAACATCTATGACTGGAGCCACGG 1513  
Db 541 TCATCTTAAACTTGCACATCCCAGGATTTAAACCAACCAACGTCTATGACTGGAGCCACAG 600  
QY 1514 GTTCTGAACCTAGGACAATAACATTTGTCCTTTGA 1547  
Db 601 GTTCGGAACCTGGGAAGGATAACATTTGTGTTGA 634

RESULT 10  
BM771726  
LOCUS  
DEFINITION K-EST0055671 S7SNU719s1 Homo sapiens cDNA clone S7SNU719s1-19-F04  
5', mRNA sequence.  
ACCESSION BM771726  
VERSION BM771726.1 GI:19101341  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 515)  
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished (2002)  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 19 row: F column: 04  
High quality sequence stop: 515.  
Location/Qualifiers  
1. 515  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="S7SNU719s1-19-F04"  
/clone\_lib="S7SNU719s1"  
/sex="M"  
/tissue\_type="Stomach"  
/cell\_type="Epithelial"  
/cell\_line="SNU-719"  
/lab\_host="Top10F"  
/note="Organ: Stomach; Vector: pcNS; Site\_1: EcoRI;  
Site\_2: NotI; The poly (A)+ RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then decapped  
with tabacco acid pyrophosphatase (TAP). The decapped  
intact mRNA was ligated with DNA-RNA linker including EcoR.  
I site by treatment of T4 RNA ligase and the first strand

cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library. After analyzing and sequencing about 2,000 - 3,000 colonies in original cDNA library, the abundant cDNAs were selected and amplified by PCR reaction using vector region primer including T7 promoter as 5' primer and N(dn)14 as 3' primer. The PCR products were used as template for synthesis of biotinylated single stranded RNA by in vitro transcription reaction. The synthesized RNA probes were hybridized with antisense single stranded cDNAs prepared from original library and incubated with avidin-gel. After removing DNA-RNA hybrids by centrifuge, the subtracted cDNA libraries were constructed by transformation of the remaining DNA into competent cells E. coli Top10F' with electroporation method."

BASE COUNT 150 a 119 c 114 g 132 t  
ORIGIN  
Query Match 15.0%; Score 507.4; DB 14; Length 515;  
Best Local Similarity 99.8%; Pred. No. 1.3e-91;  
Matches 508; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1717 GGTCAAGATAATAGACGGTTCATCAATGACATGGTGAAGATTTATTCTATCAGACCCACT 1776  
Db 7 GGTCAAGATAATAGACGGTTCATCAATGACATGGTGAAGATTTATTCTATCAGACCCACT 66  
QY 1777 AATGCAGTTGATGGGGTGGCGTCTCATGCCGTGCGCTGTCGCTCGGTTCTGAACAGTCG 1836  
Db 67 AATGCAGTTGATGGGGTGGCGTCTCATGCCGTGCGCTGTCGCTCGGTTCTGAACAGTCG 126  
QY 1837 GGTTCATCGTGTGCCCTGCCCTCCAGGCCACTACATTTGAGAAAGAAACCAACCAGTGC 1896  
Db 127 GGTTCATCGTGTGCCCTGCCCTCCAGGCCACTACATTTGAGAAAGAAACCAACCAGTGC 186  
QY 1897 AAGGAATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGAGGCTTGT 1956  
Db 187 AAGGAATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGAGGCTTGT 246  
QY 1957 ATTCCATCGGGCCCTGGAGTAAAAACAATCAGGACCATTTCGGTTTGCCTATAGTACTGC 2016  
Db 247 ATTCCATCGGGCCCTGGAGTAAAAACAATCAGGACCATTTCGGTTTGCCTATAGTACTGC 306  
QY 2017 TTTTCTTACCATGAAAAAGAAAANTCAGATTTTCGACTATGACTTTAGCAACCTCAGCAGT 2076  
Db 307 TTTTCTTACCATGAAAAAGAAAANTCAGAGTTTGCATATGACTTTAGCAACCTCAGCAGT 366  
QY 2077 GTGGGCTCATTAATGAATGGCCCCCAGCTTCACCTCCAAAGGAACAAAATACTTCCATTTC 2136  
Db 367 GTGGGCTCATTAATGAATGGCCCCCAGCTTCACCTCCAAAGGAACAAAATACTTCCATTTC 426  
QY 2137 TTCAATATCAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATA 2196  
Db 427 TTCAATATCAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATA 486  
QY 2197 ACAGACTTTACAGTAAAGAAATAGTGGC 2225  
Db 487 ACAGACTTTACAGTAAAGAAATAGTGGC 515

RESULT 11  
BG724280  
LOCUS  
DEFINITION BG724280 602697903F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:4829929 5',  
mRNA sequence.  
ACCESSION BG724280

VERSION BG724280.1 GI:14003454  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE 1 (bases 1 to 625)  
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.  
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10750 row: e column: 02  
High quality sequence stop: 623.  
Location/Qualifiers  
1. 625  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4829929"  
/clone\_lib="NIH\_MGC\_97"  
/lab\_host="DH10B"  
/note="Organ: testis; Vector: pBluescriptR (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgaag  
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
size-selected for average insert size 2.2 kb and  
normalized to R0F 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NHGRI), National  
Institutes of Health). Note: this is a NIH\_MGC Library."  
BASE COUNT 172 a 134 c 156 g 163 t  
ORIGIN

Query Match 14.7%; Score 498.4; DB 12; Length 625;  
Best Local Similarity 99.8%; Pred. No. 8.1e-90;  
Matches 497; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1393 GGTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTCTGACAATGATTAC 1452  
Db 115 GGTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTCTGACAATGATTAC 174  
QY 1453 CTGATCTTAACTGCATATCCAGGATTAAACCACCACCAACATCTATGACTGGAGCCACG 1512  
Db 175 CTGATCTTAACTGCATATCCAGGATTAAACCACCACCAACATCTATGACTGGAGCCACG 234  
QY 1513 GGTCTGAACTAGGAAGAATAACATTTGTCTTTGAGACCCCTCTGTTGAGCTGACTGTGTT 1572  
Db 235 GGTCTGAACTAGGAAGAATAACATTTGTCTTTGAGACCCCTCTGTTGAGCTGACTGTGTT 294  
QY 1573 TTGTACTTTCATGGTGATATTAAAGAAAAGTACAAATGTGGTAGATCGTGGGGTGA 1632  
Db 295 TCGTACTTTCATGGTGATATTAAAGAAAAGTACAAATGTGGTAGATCGTGGGGTGA 354  
QY 1633 ACCAAGAAAAACAAGCTTACACCCATATCATCTTCAAGAATGCAACTTTTACATTTACA 1692  
Db 355 ACCAAGAAAAACAAGCTTACACCCATATCATCTTCAAGAATGCAACTTTTACATTTACA 414  
QY 1693 TGGGCATTCCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTTCATCAATGACATGGTG 1752  
Db 415 TGGGCATTCCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTTCATCAATGACATGGTG 474  
QY 1753 AAGATTATTCTATCAGCCCACTAATGCAGTTGATGGGGTGGCGTCTCATGCCGTGCC 1812  
Db 475 AAGATTATTCTATCAGCCCACTAATGCAGTTGATGGGGTGGCGTCTCATGCCGTGCC 534

QY 1813 TGTGCCCTCGGTTCTGAACAGTCGGGTTTCATCGTGTGCCCTGCCCTCCAGGCCACTAC 1872  
Db 535 TGTGCCCTCGGTTCTGAACAGTCGGGTTTCATCGTGTGCCCTGCCCTCCAGGCCACTAC 594  
QY 1873 ATTGAGAAAAGAAACCAAC 1890  
Db 595 ATTGAGAAAAGAAACCAAC 612  
RESULT 12  
BE237235  
LOCUS BE237235 541 bp mRNA linear EST 25-APR-2001  
DEFINITION 146538 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BE237235  
VERSION BE237235.1 GI:9022025  
KEYWORDS EST.  
SOURCE COW.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 541)  
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett  
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,  
Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and  
Keele,J.W.  
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle  
JOURNAL Genome Res. 11 (4), 626-630 (2001)  
MEDLINE 21180013  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -minmatch 12 options.  
PCR Primers  
FORWARD: AGGAAACAGCTATGACCAT  
BACKWARD: GTTTTCCAGTCACGACG  
Plate: 46 row: K column: 21  
Seq primer: ATTTAGGTGACACTATAG.  
Location/Qualifiers  
1. 541  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 4BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site\_1: NotI; Site\_2: SalI;  
Library made from pooled tissue from day 20 and day 40  
embryos."  
BASE COUNT 155 a 104 c 134 g 148 t  
ORIGIN  
Query Match 14.5%; Score 493; DB 10; Length 541;  
Best Local Similarity 94.5%; Pred. No. 1e-88;  
Matches 511; Conservative 0; Mismatches 30; Indels 0; Gaps 0;  
QY 1161 TGGAGAGAAGAAGGATTGTCCGCCTTGCAACCCCTGGATTATTATAACAATGGATCTTC 1220  
Db 1 TGGAGAGAAGAAGGATTGTCCACCTTGCAACCCCTGGTTTATTATAACAATGGATCTTC 60  
QY 1221 TTGCCATCCCTGTCCTCCTCGGAACATTTTCAGATGGAACCAAGAATGTAGACCATGTCC 1280  
Db 61 CTGCCATCCCTGCCCTCCTCGGAACATTTTCGGATGGAACAAAGGAATGTAGATCATGTCC 120  
QY 1281 AGCAGGAACGGAGCCTGCACCTTGGCTTTGGAATATAAATGGTGAATGCCTTCCTGGCAA 1340  
Db 121 AGCAGGAACGGAGCCTGCACCTTGGGTTTGAATATAAATGGTGAATGTACTTCCTGGCAA 180

QY 1341 CATGAAAACTTCCTGCTTCAATGTTGGGAATTCAAAGTGCATGGAATGAATGTTGGGA 1400  
|||||  
Db 181 CATGAAAACTTCCTGTTTCAATGTTGGGAATTCAAAGTGTGACGGAATGAATGTTGGGA 240  
QY 1401 GGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTTGACATGATTACCTGATCTT 1460  
|||||  
Db 241 GGTGGCTGGAGATCACATCCAGAGTGGGGCTGGAGGTTCCGATATGATTACCTGATCTT 300  
QY 1461 AACTTGCATATCCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGGTTCTGA 1520  
|||||  
Db 301 AACTTGCATATCCCAGGATTTAAACCACCAACATCGATGACTGGAGCCATGGGTTCTGA 360  
QY 1521 ACTAGGAAGAATAACATTTGTCTTTTGAGACCCTCTGTTTCAGCTGACTGTGTTTGTACTT 1580  
|||||  
Db 361 ACTGGGAAGAATAACATTTGTTTGTAGACCCTCTGTTTCAGCTGACTGTGTTTGTACTT 420  
QY 1581 CATGGTGGATATTAAGAAAAAGTACAAATGTGGTAGAATCGTGGGTGGAACCAAAGA 1640  
|||||  
Db 421 CATGGTGGATATTAAGAAAAAGTACGAATGTGGTGAGTCAATGGGTGGAACCAAAGA 480  
QY 1641 AAAACAAGCTTACACCCATATCATCTTCAAGAANTGCAACTTTTACATTTACATGGGCATT 1700  
|||||  
Db 481 AAAACAAGCTTATACCCATCATCTTCAAGAANTGCCACATTTACATTTACATGGGCATT 540  
QY 1701 C 1701  
Db 541 C 541

RESULT 13  
BQ772508/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

713 bp mRNA linear EST 26-JUL-2002  
UI-H-FEO-bbq-b-01-0-UI.s1 NCI\_CGAP\_FEO Homo sapiens cDNA clone  
UI-H-FEO-bbq-b-01-0-UI 3', mRNA sequence.  
BQ772508  
BQ772508.1 GI:21980984  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 713)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)  
Tissue Procurement: James Martin  
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)  
The following repetitive elements were found in this cDNA  
sequence: 1-40, >AT-rich#Low complexity (matched complement)  
Seq primer: M13 FORWARD  
POLYA=Yes.

FEATURES  
source  
1. 713  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-H-FEO-bbq-b-01-0-UI"  
/clone\_lib="NCI\_CGAP\_FEO"  
/tissue\_type="Chondrosarcoma Cell line"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT7T3-Pac (Pharmacia) with a modified  
polylinker; Site\_1: EcoR I; Site\_2: Not I; NCI\_CGAP\_FEO is  
a cDNA library containing the following tissue(s): a pool  
of 3 chondrosarcoma cell lines (grade 2) The library was  
constructed according to Bonaldo, Lennon and Soares,

Genome Research, 6:791-806, 1996. First strand cDNA  
synthesis was primed with an oligo-dT primer containing a  
Not I site. Double stranded cDNA was ligated to an EcoR I  
adaptor, digested with Not I, and cloned directionally  
into pT7T3-Pac vector. The oligonucleotide used to prime  
the synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
(dT)18 tail. The sequence tag for this library is  
CGCTACGGAC. The cell lines was provided by Dr James Martin  
of University of Iowa.  
TAG\_LIB=UI-H-FEO  
TAG\_TISSUE=Human grade 2 chondrosarcoma cell line pool  
TAG\_SEQ=CGCTACGGAC"

BASE COUNT 193 a 136 c 120 g 258 t 6 others  
ORIGIN  
Query Match 14.5%; Score 492.8; DB 14; Length 713;  
Best Local Similarity 97.4%; Pred. No. 1e-88;  
Matches 532; Conservative 0; Mismatches 9; Indels 5; Gaps 3;  
QY 2850 GAATCAAAAGAAAAAGAGACCAATTTTGAATCTGTTCAACTGAAAAACCTCAAGATCCCCA 2909  
| | | | |  
Db 547 GCAACCAAGGAAAAAGAGACCAATTTTGAATCTGTTCAACTGAAAAACCTCAAGATCCCCA 488  
QY 2910 AATATATGAAGAGACAGTGTCTGAGCCTTGAGACTAATGAACAAAGAAACCTGCTCTAGT 2969  
| | | | |  
Db 487 AATATATGAAGAGACAGTGTCTGAGCCTTGAGACTAATGAACAAAGAAACCTGCTCTAGT 428  
QY 2970 TTTACAGGACCATATTTTAGGGTCTGTCTCCTCATACCTGTGCACATTTGGTGATCTCACAGAG 3029  
| | | | |  
Db 427 TTTACAGGACCATATTTTAGGGTCTGTCTCCTCATACCTGTGCACATTTGGTGATCTCACAGAG 368  
QY 3030 GAGGGCCATGCCGCTGAAAAGGGAGGAGATTGAAACATTTGATTGCCCTTATCACATGGT 3089  
| | | | |  
Db 367 GAGGGCCATGCCGCTGAAAAGGGAGGAGATTGAAACATTTGATTGCCCTTATCACATGGT 308  
QY 3090 CAAGTACCTTGCCAAATAAAGGAAAGCAAAATGATTTGGGTCTCAACTGAAGATGAAGCTC 3149  
| | | | |  
Db 307 CAAGTACCTTGCCAAATAAAGGAAAGCAAAATGATTTGGGTCTCAACTGAAGATGAAGCTC 248  
QY 3150 AACTCAGGAAGAGATTTATCTGTATATACACATAACTGAAACCAAGTTTAAAGCCACCA 3209  
| | | | |  
Db 247 AACTCAGGAAGAGATTTATCTGTATATACACATAACTGAAACCAAGTTTAAAGCCACCA 188  
QY 3210 ATGCACTGCTGATGCATGCCATATAATTAATGGGTAACCTTTTATTTATGATGCTTAC 3269  
| | | | |  
Db 187 ATGCACTGCTGATGCATGCCATATAATTAATGGGTAACCTTTTATTTATGATGCTTAC 128  
QY 3270 ATAA-CAAAGTGTGATTTGGAAGGCACA-TGTGAGCATATGCATTTAT---GATCCAAATTTA 3324  
| | | | |  
Db 127 ATAACCAAGTGTGATTTGGAAGGCACACANTGTGAGCATATGCATTNNATGATCCCCAATTTA 68  
QY 3325 TGTTTTCTTTTGTATTATATTTTGGGGAAATTTAAATTTTAAAGGTAAAAAATAAAAA 3384  
| | | | |  
Db 67 TGTTTTCTTTTGTATTATATTTTGGGGAAATTTAAATTTTAAAGGTAAAAAATAAAAA 8  
QY 3385 AAAAAA 3390  
| | | | |  
Db 7 AAAAAA 2

RESULT 14  
W60843  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

W60843  
zd26fl0.r1 Soares\_fetal\_heart\_NbHH19W Homo sapiens cDNA clone  
IMAGE:341803 5', mRNA sequence.  
W60843  
W60843.1 GI:1367601  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.



```
REFERENCE 1 (bases 1 to 557)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohldmann,P. and Wilson,R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1674 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 391.
Location/Qualifiers
1. .557
/organism="Homo sapiens"
/db_xref="GDB:1267178"
/db_xref="taxon:9606"
/clone="IMAGE:341803"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGGCATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHL19W."

BASE COUNT 180 a 117 c 107 g 152 t 1 others
ORIGIN

Query Match 14.2%; Score 482.2; DB 14; Length 557;
Best Local Similarity 98.1%; Pred. No. 1.5e-86;
Matches 550; Conservative 0; Mismatches 4; Indels 6; Gaps 6;

QY 1916 CCTACCTGCCATACATCAGGTCATGGCAAAGAGCGCTTGATTCCTCATCGGGCCTGGGA 1975
|||||
Db 1 CCTACCTGCCATACATCAGGTCATGGCAAAGAGCGCTTGATTCCTCATCGGGCCTGGGA 60

QY 1976 GTAAAAACAATCAGGACCATTGCGGTTTGCTATAGTGACTGCTTTTCTACCATGAAAAAG 2035
|||||
Db 61 GTAAAAACAATCAGGACCATTGCGGTTTGCTATAGTGACTGCTTTTCTACCATGAAAAAG 120

QY 2036 AAAATCAGATTTTGCACATATGACTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATG 2095
|||||
Db 121 AAAATCAGAGTTTGCACATATGACTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATG 180

QY 2096 GCCCCAGCTTCACCTCCAAAGGAACAAAATACTTCCATTTCTCAATATCAGTTTATGTG 2155
|||||
Db 181 GCCCCAGCTTCACCTCCAAAGGAACAAAATACTTCCATTTCTCAATATCAGTTTATGTG 240

QY 2156 GGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTTTACA-GTAAAA 2214
|||||
Db 241 GGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTTTACACGTA AAA 300

QY 2215 GAAATAGTGGCAGGGTCAGATGATTACACAAATTTGGTAGGGGCATTTGTATGCCAGTCA 2274
|||||
Db 301 GAAATAGTGGCAGGGTCAGATGATTACACAAATTTGGTAGGGGCATTTGTATG-CAGTCA 359

QY 2275 ACAATTATTCTTCTGAAAGTAA-GGGTTTCCGAGCAGCCTTATCATCACAAT-CCATCA 2332
|||||
```

```
Db 360 ACAATATTCTCTGAAAGTAAGGGGTTTCCGAGCAGCGCTTATCATCACAATCCCATCA 419
QY 2333 TTCTGGCAGATACATTTCATAGGAGTCACAGTTTGAAACCCACATTTGAAAAATATTAATA 2392
|||||
Db 420 TTCTGGCAGATACATTTCATAGGAGTCACAGTTTGAAACCCACATTTG-AAAATATTAATA 477
QY 2393 AAGAAATATGTTCCCGAGTTCCAACAAGCCAAATACCAGATGTGCATTTCTTTTATAAGT 2452
|||||
Db 478 AAGAAATATGTTCCCGAGTTCCAACAAGCCAAATACCNGATGTGCATTTCTTTTATAAGT 537
QY 2453 CTCTACAGCAACAACATCT 2472
|||||
Db 538 CTCTAAAGCAACAACATCT 557

RESULT 15
W60844/c
LOCUS W60844
DEFINITION zG26f10.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
IMAGE:341803 3', mRNA sequence.
ACCESSION W60844
VERSION W60844.1 GI:1367602
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 576)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohldmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1674 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 336.
Location/Qualifiers
1. .576
/organism="Homo sapiens"
/db_xref="GDB:1267178"
/db_xref="taxon:9606"
/clone="IMAGE:341803"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGGCATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHL19W."

BASE COUNT 171 a 112 c 108 g 182 t 3 others
ORIGIN

Query Match 14.2%; Score 482.2; DB 14; Length 576;
Best Local Similarity 97.1%; Pred. No. 1.5e-86;
Matches 501; Conservative 0; Mismatches 14; Indels 1; Gaps 1;
```

Qy 2860 AAAAAGAGACCATTTTGAATCTGTT-CAACTGAAAAACCTCAAGATCCCCAAATATATGA 2918  
Db 516 AAAAGGAGACCCCATTTTGAATCTGTTCCAACTGAAACCTCAAGATCCCCAAATATATGA 457

Qy 2919 AGAGACAGTGTCTGTAGCCCTTGAGACTTAATGAACAAAGAAACCTGCTCTAGTTTACAGGA 2978  
Db 456 AGAGACAGTGTGTAGCCCTTGAGACTAATGANCAAGGAAACCTGCTCTAGTTTACAGGA 397

Qy 2979 CCATATTTTAGGGTCTGTCCCTCATACCTGTGCACATTGGTGATCTCACAGAGGAGGGCCAT 3038  
Db 396 CCATATTTTAGGGTCTGTCCCTCATACCTGTGCACATTGGTGATCTCACAGAGGAGGGCCAT 337

Qy 3039 GCCGCTGAAAAGGGAAGGAGATTGAAACATTTTGATTGCCCTTATCACATGGTCAAGTACCT 3098  
Db 336 GCCGCTGAAAAGGGAAGGAGATTGAAACATTTTGATTGCCCTTATCACATGGTCAAGTACCT 277

Qy 3099 TGCCAAATAAGGAAAGCAAAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGA 3158  
Db 276 TGCCAAATAAGGAAAGCAAAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGA 217

Qy 3159 AGAGATTATCTGTATATACACATAACTGAAACCAGTTTAAAGCCCAACCAATGCACCTGC 3218  
Db 216 AGAGATTATCTGTATATACACATAACTGAAACCAGTTTAAAGCCCAACCAATGCACCTGC 157

Qy 3219 TGATGCATGCCATATAATTAATGGGTAACCTTTTATCTCTTTATGATGTCTACATAACAAGT 3278  
Db 156 TGATGCATGCCATATAATTAATGGGTAACCTTTTATCTCTTTATGATGTCTACATAACAAGT 97

Qy 3279 GTGATTTGGAAGGCACATGTGAGCATATGCATTATGATCCCAATTTATGTTTTCCTTTGT 3338  
Db 96 GTGATTTGGAAGGCACATGTGAGCATATGCATTATGATCCCAATTTATGTTTTCCTTTGT 37

Qy 3339 TTATATTTTGGGGGAAAAATTAAAAATTTTAAAGGTA 3374  
Db 36 TTATATTTTGGGGGAAAAATTAAAAATTTTAAAGGTA 1

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: May 12, 2003, 05:40:13 ; Search time 521.238 Seconds  
(without alignments)  
8082.236 Million cell updates/sec

Title: US-10-073-333A-1  
Perfect score: 3390  
Sequence: 1 atgtgttccgcgcgcgggg.....ggtaaaaaaaaaaaaaaaa 3390

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	3390	100.0	3390	9	US-10-140-164-1	Sequence 1, Appli
2	3214	94.8	3556	9	US-10-140-164-3	Sequence 3, Appli
3	1499.6	44.2	1737	9	US-10-002-050-19	Sequence 19, Appl
4	1499.6	44.2	1737	9	US-10-002-304-19	Sequence 19, Appl
5	1499.6	44.2	1737	12	US-10-003-152-19	Sequence 19, Appl
6	1182.6	34.9	1508	9	US-10-002-050-9	Sequence 9, Appli
7	1182.6	34.9	1508	9	US-10-002-304-9	Sequence 9, Appli
8	1182.6	34.9	1508	12	US-10-003-152-9	Sequence 9, Appli
9	901	26.6	3501	9	US-10-028-072-37	Sequence 37, Appl
10	901	26.6	3501	9	US-10-121-049-37	Sequence 37, Appl
11	901	26.6	3501	9	US-10-123-904-37	Sequence 37, Appl
12	901	26.6	3501	9	US-10-140-470-37	Sequence 37, Appl
13	901	26.6	3501	9	US-10-175-746-37	Sequence 37, Appl
14	901	26.6	3501	9	US-10-176-918-37	Sequence 37, Appl
15	901	26.6	3501	9	US-10-176-921-37	Sequence 37, Appl
16	901	26.6	3501	9	US-10-137-865-37	Sequence 37, Appl
17	901	26.6	3501	9	US-10-140-474-37	Sequence 37, Appl
18	901	26.6	3501	9	US-10-142-431-37	Sequence 37, Appl
19	901	26.6	3501	9	US-10-143-114-37	Sequence 37, Appl

20	901	26.6	3501	9	US-10-140-002-37	Sequence 37, Appl
21	901	26.6	3501	9	US-10-142-419-37	Sequence 37, Appl
22	901	26.6	3501	9	US-10-123-262-37	Sequence 37, Appl
23	901	26.6	3501	9	US-10-142-423-37	Sequence 37, Appl
24	901	26.6	3501	9	US-10-121-050-37	Sequence 37, Appl
25	901	26.6	3501	9	US-10-141-755-37	Sequence 37, Appl
26	901	26.6	3501	9	US-10-143-032-37	Sequence 37, Appl
27	901	26.6	3501	9	US-10-123-108-37	Sequence 37, Appl
28	901	26.6	3501	9	US-10-123-236-37	Sequence 37, Appl
29	901	26.6	3501	9	US-10-123-261-37	Sequence 37, Appl
30	901	26.6	3501	9	US-10-140-921-37	Sequence 37, Appl
31	901	26.6	3501	9	US-10-140-928-37	Sequence 37, Appl
32	901	26.6	3501	9	US-10-121-045-37	Sequence 37, Appl
33	901	26.6	3501	9	US-10-123-292-37	Sequence 37, Appl
34	901	26.6	3501	9	US-10-123-903-37	Sequence 37, Appl
35	901	26.6	3501	9	US-10-124-819-37	Sequence 37, Appl
36	901	26.6	3501	9	US-10-124-822-37	Sequence 37, Appl
37	901	26.6	3501	9	US-10-140-925-37	Sequence 37, Appl
38	901	26.6	3501	9	US-10-160-498-37	Sequence 37, Appl
39	901	26.6	3501	9	US-10-121-041-37	Sequence 37, Appl
40	901	26.6	3501	9	US-10-121-043-37	Sequence 37, Appl
41	901	26.6	3501	9	US-10-121-047-37	Sequence 37, Appl
42	901	26.6	3501	9	US-10-123-215-37	Sequence 37, Appl
43	901	26.6	3501	9	US-10-123-902-37	Sequence 37, Appl
44	901	26.6	3501	9	US-10-123-908-37	Sequence 37, Appl
45	901	26.6	3501	9	US-10-123-909-37	Sequence 37, Appl

ALIGNMENTS

RESULT 1  
US-10-140-164-1  
; Sequence 1, Application US/10140164  
; Publication No. US20030072736A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker et al.  
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16  
; FILE REFERENCE: PF514C1  
; CURRENT APPLICATION NUMBER: US/10/140,164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 3390  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2892)  
US-10-140-164-1

Query Match 100.0%; Score 3390; DB 9; Length 3390;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGCTGTTCGGCGCGGGGGCGGTACGGGCGAGGGGCTGGGGCGCGCGGAGGCT 60



Db 1 ATGCTGTTCCGCGCCGGGGCCGGTACGGGGCAGGGGCTGGGGCGCGCGGAGGCT 60  
QY 61 CCCGCGCGGGCGCTGCGCGCCCTGGAGCCCGCGCTGGATTTGCTGCTGGGCGCTCGCC 120  
Db 61 CCCGCGCGGGCGCTGCGCGCCCTGGAGCCCGCGCTGGATTTGCTGCTGGGCGCTCGCC 120  
QY 121 GGCTGCCAGGGCGGCTGGGACCTGCGCCCTCCTCCTCCAGCCCGCGCTTCCTCCT 180  
Db 121 GGCTGCCAGGGCGGCTGGGACCTGCGCCCTCCTCCTCCAGCCCGCGCTTCCTCCT 180  
QY 181 TGCCAGGAGAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGG 240  
Db 181 TGCCAGGAGAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGG 240  
QY 241 AGAGTTGCCATTCCAAATTCGCAGTGGACTGCTCTGGCCTGCCCTGACCCAGTGAGAGGC 300  
Db 241 AGAGTTGCCATTCCAAATTCGCAGTGGACTGCTCTGGCCTGCCCTGACCCAGTGAGAGGC 300  
QY 301 AAAGAATGCACCTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGC 360  
Db 301 AAAGAATGCACCTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGC 360  
QY 361 AGTAAGTGTGGTGAAGGCACCTATTCTTGGGCAGTGGCATCAAAATTTGATGAATGGGAT 420  
Db 361 AGTAAGTGTGGTGAAGGCACCTATTCTTGGGCAGTGGCATCAAAATTTGATGAATGGGAT 420  
QY 421 GAATTGCCGGCAGGATTTTCTAAGTCGCAACTCTTCTTGGATCCCTCGTGGAACTACATAGAA 540  
Db 421 GAATTGCCGGCAGGATTTTCTAAGTCGCAACTCTTCTTGGATCCCTCGTGGAACTACATAGAA 540  
QY 481 GACAGCAGGCCAGACGGCTGTAACTCTTCTTGGATCCCTCGTGGAACTACATAGAA 540  
Db 481 GACAGCAGGCCAGACGGCTGTAACTCTTCTTGGATCCCTCGTGGAACTACATAGAA 540  
QY 541 TCTAATCGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGC 600  
Db 541 TCTAATCGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGC 600  
QY 601 TATGCTCTCTTTGAGTACCAGTATGCGACAACAACATCTTCTTTGAGTTCTTTATTCAA 660  
Db 601 TATGCTCTCTTTGAGTACCAGTATGCGACAACAACATCTTCTTTGAGTTCTTTATTCAA 660  
QY 661 AATGATCAGTCCAGGAGATGGACACCACTGACAAGTGGTAAACCTTACAGACAAT 720  
Db 661 AATGATCAGTCCAGGAGATGGACACCACTGACAAGTGGTAAACCTTACAGACAAT 720  
QY 721 GGAGAATGGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATCTACTGGAGA 780  
Db 721 GGAGAATGGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATCTACTGGAGA 780  
QY 781 ACTACAGGCATCCTTATGGGTTCTAAGCGGTCAGCCCTGTGCTGGTAAANAATATCACA 840  
Db 781 ACTACAGGCATCCTTATGGGTTCTAAGCGGTCAGCCCTGTGCTGGTAAANAATATCACA 840  
QY 841 ATTGAAGGGTGGGCTACACATCAGAAATGTTTCTTGAAGCCAGGCACATTCAGCAAC 900  
Db 841 ATTGAAGGGTGGGCTACACATCAGAAATGTTTCTTGAAGCCAGGCACATTCAGCAAC 900  
QY 901 AAACCAAGTTTCAATCAACTGCCAGGTGTGCCAGAAACACCTATTCTGAGAAAGGAGCC 960  
Db 901 AAACCAAGTTTCAATCAACTGCCAGGTGTGCCAGAAACACCTATTCTGAGAAAGGAGCC 960  
QY 961 AAAGAATGTATAAGGTGTAAAGACGACTCTCAATTTTCAGGATCCAGTGATGATAGAG 1020  
Db 961 AAAGAATGTATAAGGTGTAAAGACGACTCTCAATTTTCAGGATCCAGTGATGATAGAG 1020  
QY 1021 CGCCCTCCCTGTACCACAAAGACTATTTCCAGATCCATACTCCATGTGATGAAGAAGGA 1080  
Db 1021 CGCCCTCCCTGTACCACAAAGACTATTTCCAGATCCATACTCCATGTGATGAAGAAGGA 1080  
QY 1081 AAGACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGAT 1140  
Db 1081 AAGACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGAT 1140

Db 1081 AAGACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGAT 1140  
QY 1141 GCTATTAGATTGCCCCCTTCTGGAGAGAAGAAGATTTGTCGCCCTTGCAACCCCTGGATTT 1200  
Db 1141 GCTATTAGATTGCCCCCTTCTGGAGAGAAGAAGATTTGTCGCCCTTGCAACCCCTGGATTT 1200  
QY 1201 TATAACAATGGATCATCTTCTTGCCATCCCTGTCTCCTGGAACNATTTTCAGATGGAACC 1260  
Db 1201 TATAACAATGGATCATCTTCTTGCCATCCCTGTCTCCTGGAACATTTTCAGATGGAACC 1260  
QY 1261 AAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCCTGCACCTTGGCTTTGAATATAAATGG 1320  
Db 1261 AAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCCTGCACCTTGGCTTTGAATATAAATGG 1320  
QY 1321 TGGAATGTCTTCTTGGAACATGAAAACCTTCTTCAATGTTGGGAATTCAAAGTGC 1380  
Db 1321 TGGAATGTCTTCTTGGAACATGAAAACCTTCTTCAATGTTGGGAATTCAAAGTGC 1380  
QY 1381 GATGGAATGAATGGTTGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGTTCCT 1440  
Db 1381 GATGGAATGAATGGTTGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGTTCCT 1440  
QY 1441 GACAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAACCACCAACATCTATG 1500  
Db 1441 GACAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAACCACCAACATCTATG 1500  
QY 1501 ACTGGAGCCACGGGTTCTGAACTAGGAAGAATAACATTTCTCTTTGAGACCCCTCTGTCA 1560  
Db 1501 ACTGGAGCCACGGGTTCTGAACTAGGAAGAATAACATTTCTCTTTGAGACCCCTCTGTCA 1560  
QY 1561 GCTGACTGTGTTTTGTACTTTCATGGTGGATATTAAATAGAAAAGTACAAATGTGGTAGAA 1620  
Db 1561 GCTGACTGTGTTTTGTACTTTCATGGTGGATATTAAATAGAAAAGTACAAATGTGGTAGAA 1620  
QY 1621 TCGTGGGTTGGAACCAAGAAACAAAGCTTACACCCATATCATCTTCAAGAATGCAACT 1680  
Db 1621 TCGTGGGTTGGAACCAAGAAACAAAGCTTACACCCATATCATCTTCAAGAATGCAACT 1680  
QY 1681 TTTTACATTTACATGGGCTTCCAGAGAACTAATCAGGGTCAAGATTAATAGACGGTTTCATC 1740  
Db 1681 TTTTACATTTACATGGGCTTCCAGAGAACTAATCAGGGTCAAGATTAATAGACGGTTTCATC 1740  
QY 1741 AATGACATGGTGAAGATTTATTTATCACAGCCACTAATGCAGTTGATGGGTGGCGTCC 1800  
Db 1741 AATGACATGGTGAAGATTTATTTATCACAGCCACTAATGCAGTTGATGGGTGGCGTCC 1800  
QY 1801 TCATGCCGTCCTGTGCCCTCGGTTCTGAACAGTCCGGTTTCATGCTGTCTCCCTGCCCT 1860  
Db 1801 TCATGCCGTCCTGTGCCCTCGGTTCTGAACAGTCCGGTTTCATGCTGTCTCCCTGCCCT 1860  
QY 1861 CCAGGCCACTACATTGAGAAAGAAACCAACAGTGAAGGAATGTCCACCTGACACCTAC 1920  
Db 1861 CCAGGCCACTACATTGAGAAAGAAACCAACAGTGAAGGAATGTCCACCTGACACCTAC 1920  
QY 1921 CTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTCATGCGGGCTGGGAGTAAA 1980  
Db 1921 CTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTCATGCGGGCTGGGAGTAAA 1980  
QY 1981 AACAATCAGGACCATTCGGTTTGTCTATAGTACTGCTTTTCTACCAATGAAGAAAT 2040  
Db 1981 AACAATCAGGACCATTCGGTTTGTCTATAGTACTGCTTTTCTACCAATGAAGAAAT 2040  
QY 2041 CAGATTTTGCACATGACTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGGCCCC 2100  
Db 2041 CAGATTTTGCACATGACTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGGCCCC 2100  
QY 2101 AGCTTCACCTCCAAAGGAACAAATACITTCATTTCTCAATATCAGTTTATGTGGGCAT 2160  
Db 2101 AGCTTCACCTCCAAAGGAACAAATACITTCATTTCTCAATATCAGTTTATGTGGGCAT 2160  
QY 2161 GAGGGGAAGAAGTGGCTCTCTGTACCAACAAATATTAACAGACTTTTACAGTAAAGAAATA 2220  
Db 2161 GAGGGGAAGAAGTGGCTCTCTGTACCAACAAATATTAACAGACTTTTACAGTAAAGAAATA 2220

Qy	2221	GTGCAGGTCAGATGATTACACAAATTTGGTAGGGCATTTGTATGCCAGTCAACAATT	2280
Db	2221		
Qy	2281	ATTCCTTCGAAAGTAAGGGTTTCCGAGCAGCCTTATCATCACAATCCATCTCTGGCA	2340
Db	2281		
Qy	2341	GATACATTCATAGGAGTCACAGTTGAAACCACATTTGAAAAATATTAAATAAAAAGAGAT	2400
Db	2341		
Qy	2401	ATGTTCCAGTTCCAACAAGCCAAATACCAGATGTGCATTTCTTTATAAGTCTTCTACA	2460
Db	2401		
Qy	2461	GCAACAACATCTTGTTAATAGCGCATCAACTGCTGTGAAAAAGAGGTGTAATCCTACT	2520
Db	2461		
Qy	2521	AAATCTGAGCAGGAGTGATTTAGTCCCCAGCAAGTGCCCCAGCAGGTACCTGTGATGGG	2580
Db	2521		
Qy	2581	TGTACGTTCTATTTCCTGTGGGAGAGTGCTGAAGCTTGCCCTCTGTGTACGGAGCATGAC	2640
Db	2581		
Qy	2641	TTCCATGAGATTGAGGGAGCCTGCAAGAGAGGATTTCAGGAACCTTGTTATGTGTGGAAT	2700
Db	2641		
Qy	2701	GAACCTAAATGGTGCAATTAAGGAATTTCTTTGCCTGAGAAAAAGTTGGCAACCTGTGAA	2760
Db	2701		
Qy	2761	ACGGTTGACTTTTGGCTGAAGGTGGAGCCGGTGTGGGAGCTTTTACTGCCGTTTGTCTG	2820
Db	2761		
Qy	2821	GTGGCTGTGACCTGTACTTCTGAAAAAAGAAATCAAAAGAAAAAGAACCATTTTGAAT	2880
Db	2821		
Qy	2881	CTGTTCAACTGAAACCTCAAGATCCCCCAATATATGAAGACACAGTGTAGCCTTGA	2940
Db	2881		
Qy	2941	GACTAATGAACAAAGAAACCTGCTCTAGTTTACAGGACCATATTTAGGGTCTGTCCTC	3000
Db	2941		
Qy	3001	ATACCTGTACATTGGTGATCTCACAGAGGAGGGCCATGCCGTGAAAAAGGGAAGAGAT	3060
Db	3001		
Qy	3061	TGAAACATTTGATTGCCCTTATCACATGGTCAAGTACCTTGCCAAATAAAGGAAAGCAAT	3120
Db	3061		
Qy	3121	GATTTGGTCTCAACTGAAGATGAAGCTCAACTCAGGAAGAGATTTATCTGTATATACAC	3180
Db	3121		
Qy	3181	ATAACTGAAACCAAGTTTAAGCCCAACCAATGCACCTGCATGCATGCCATATAATTAAT	3240
Db	3181		
Qy	3241	GGTAACCTTTATCTTTATGATGTCTACATAACAAGTGTGATTTGGAAGGCACATGTGA	3300
Db	3241		

Qy	3301	GCATATGCATATGATCCAAATTTATGTTTCTTTCTTTTATATTTGGGAAAAATAAA	3360
Db	3301		
Qy	3361	ATTTTAAAGTAAAAAATAAAAAA	3390
Db	3361		
RESULT 2			
US-10-140-164-3			
; Sequence 3, Application US/10140164			
; Publication No. US20030072736A1			
; GENERAL INFORMATION:			
; APPLICANT: Baker et al.			
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16			
; FILE REFERENCE: PF514C1			
; CURRENT APPLICATION NUMBER: US/10/140,164			
; CURRENT FILING DATE: 2002-05-08			
; PRIOR APPLICATION NUMBER: 09/637,856			
; PRIOR FILING DATE: 2000-08-10			
; PRIOR APPLICATION NUMBER: 60/148,348			
; PRIOR FILING DATE: 1999-08-12			
; PRIOR APPLICATION NUMBER: 60/148,683			
; PRIOR FILING DATE: 1999-08-13			
; PRIOR APPLICATION NUMBER: 60/148,870			
; PRIOR FILING DATE: 1999-08-13			
; PRIOR APPLICATION NUMBER: 60/148,758			
; PRIOR FILING DATE: 1999-08-16			
; PRIOR APPLICATION NUMBER: 60/149,181			
; PRIOR FILING DATE: 1999-08-17			
; PRIOR APPLICATION NUMBER: 60/149,453			
; PRIOR FILING DATE: 1999-08-18			
; PRIOR APPLICATION NUMBER: 60/149,498			
; PRIOR FILING DATE: 1999-08-19			
; NUMBER OF SEQ ID NOS: 76			
; SOFTWARE: Patentin Ver. 2.1			
; SEQ ID NO 3			
; LENGTH: 3556			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (1)..(3084)			
US-10-140-164-3			

Query Match 94.8%; Score 3214; DB 9; Length 3556;  
Best Local Similarity 95.3%; Pred. No. 0;  
Matches 3390; Conservative 0; Mismatches 0; Indels 166; Gaps 1;

Qy	1	ATGCTGTCCGCGCCCGGGCCGGTACGGGGCAGGGCTGGGGCGCGCGGAGGCT	60
Db	1		
Qy	61	CCCCCGCGGGCGCTCGCCGCCCTGGAGCCCGCCCTGGATTGTGCTGGCGCTCGCC	120
Db	61		
Qy	121	GGCTGCCAGCGCGCTGGGCTGGGACCTGCCCTCCTCCTCCAGCCCGCCCTTCCCTCCT	180
Db	121	GGCTGCCAGCGCGCTGGGCTGGGACCTGCCCTCCTCCTCCAGCCCGCCCTTCCCTCCT	180
Qy	181	TGCCAGGAGAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGG	240
Db	181	TGCCAGGAGAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGG	240
Qy	241	AGAGTTGCCATTCCAAATTCTGCAGTGGACTGCTCTGGCCTGCCTGACCCAGTGAGAGGC	300
Db	241	AGAGTTGCCATTCCAAATTCTGCAGTGGACTGCTCTGGCCTGCCTGACCCAGTGAGAGGC	300
Qy	301	AAAGAAATGCACCTTCTCCTGTGCTTCTGGAGAGTATCTAGAAAATGAAGAACCCAGGTATGC	360
Db	301	AAAGAAATGCACCTTCTCCTGTGCTTCTGGAGAGTATCTAGAAAATGAAGAACCCAGGTATGC	360



QY	361	AGTAAGTGTGGTGAAGGCACCTATTCTCTTGGCAGTGGCATCAAATTTGATGAATGGGAT	420
Db	361	AGTAAGTGTGGTGAAGGCACCTATTCTCTTGGCAGTGGCATCAAATTTGATGAATGGGAT	420
QY	421	GAATGCCGGCAGGATTTTCTAACATCGCAACATTTCATGGACACTGTGGTGGGCCCTTCT	480
Db	421	GAATGCCGGCAGGATTTTCTAACATCGCAACATTTCATGGACACTGTGGTGGGCCCTTCT	480
QY	481	GACAGCAGGCCAGACGGCTGTAAACAACCTCTTGGATCCCTCGTGGAAACTACATAGAA	540
Db	481	GACAGCAGGCCAGACGGCTGTAAACAACCTCTTGGATCCCTCGTGGAAACTACATAGAA	540
QY	541	TCTAATCGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGC	600
Db	541	TCTAATCGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGC	600
QY	601	TATGCTCTCTTTGAGTACCAGTATGTCGACACAAACATCTTCTTTGAGTTCTTTATTCAA	660
Db	601	TATGCTCTCTTTGAGTACCAGTATGTCGACACAAACATCTTCTTTGAGTTCTTTATTCAA	660
QY	661	AATGATCAGTCCAGGAGATGGACACCACCTGACAAAGTGGGTAAAACTTACAGACAAT	720
Db	661	AATGATCAGTCCAGGAGATGGACACCACCTGACAAAGTGGGTAAAACTTACAGACAAT	720
QY	721	GGAGAAATGGGCTCTCATCTGTAATGCTGAATCAGGCACAAAACATACTCTACTGGAGA	780
Db	721	GGAGAAATGGGCTCTCATCTGTAATGCTGAATCAGGCACAAAACATACTCTACTGGAGA	780
QY	781	ACTACAGGCATCCTTATGGGTTCTAAGGGGGTCAAGCCTGTGCTGTAATAATATCACA	840
Db	781	ACTACAGGCATCCTTATGGGTTCTAAGGGGGTCAAGCCTGTGCTGTAATAATATCACA	840
QY	841	ATTGAAGGGTGGCGTACACATCAGAAATGTTTTCTTGAAGCCAGGCACATTCAGCAAC	900
Db	841	ATTGAAGGGTGGCGTACACATCAGAAATGTTTTCTTGAAGCCAGGCACATTCAGCAAC	900
QY	901	AAACCAGGTTCAATCAACTGCCAGGTGTGCCAGAAACACCTATTCTGAGAAAGGAGCC	960
Db	901	AAACCAGGTTCAATCAACTGCCAGGTGTGCCAGAAACACCTATTCTGAGAAAGGAGCC	960
QY	961	AAAGAAATGTATAAGGTGTAAAGACGACTCTCAATTTTCCAGATCCATCTGATGAAGAAGGA	1020
Db	961	AAAGAAATGTATAAGGTGTAAAGACGACTCTCAATTTTCCAGATCCATCTGATGAAGAAGGA	1020
QY	1021	CGCCCTCCCTGTACCCACAAAAGACTATTTCCAGATCCATCTGATGAAGAAGGA	1080
Db	1021	CGCCCTCCCTGTACCCACAAAAGACTATTTCCAGATCCATCTGATGAAGAAGGA	1080
QY	1081	AAGACACAGATAATGTACAAGTGGATAGAGCCCCAAAATCTGCCGGAGGATCTCACAGAT	1140
Db	1081	AAGACACAGATAATGTACAAGTGGATAGAGCCCCAAAATCTGCCGGAGGATCTCACAGAT	1140
QY	1141	GCTATTAGATTGCCCCCTTCTGGAGAGAAAGGATTGTCCGCCCTTGCAACCCCTGGATTT	1200
Db	1141	GCTATTAGATTGCCCCCTTCTGGAGAGAAAGGATTGTCCGCCCTTGCAACCCCTGGATTT	1200
QY	1201	TATAACAATGGATCATCTTCTGCCATCCCTGTCTCCTCGGAACATTTTCAGATGGAACC	1260
Db	1201	TATAACAATGGATCATCTTCTGCCATCCCTGTCTCCTCGGAACATTTTCAGATGGAACC	1260
QY	1261	AAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAATGG	1320
Db	1261	AAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAATGG	1320
QY	1321	TGGAATGTCTCTTCCCTGGCAACATGAAAACCTTCCCTGCTTCAATGTTGGAATTCAAAGTGC	1380
Db	1321	TGGAATGTCTCTTCCCTGGCAACATGAAAACCTTCCCTGCTTCAATGTTGGAATTCAAAGTGC	1380
QY	1381	GATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGTTCT	1440
Db	1381	GATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGTTCT	1440

2

1



|||||  
Db 2521 AAATCTGGAGCAGGAGTGATTTTCAGTCCCAGCAAGTGCCAGCAGGTACCTGTGATGGG 2580  
QY 2581 TGTAAGTTCTATTTCCCTGTGGGAGAGTGCTGAAGCTTGCCCTCTGTGTACGGAGCATGAC 2640  
Db 2581 TGTACCTTCTATTTCCCTGTGGGAGAGTGCTGAAGCTTGCCCTCTGTGTACGGAGCATGAC 2640  
QY 2641 TTCCATGAGATTGAGGGAGCCCTGCAAGAGAGGATTTTCAGGAAACCTTGTATGTGTGGAAT 2700  
Db 2641 TTCCATGAGATTGAGGGAGCCCTGCAAGAGAGGATTTTCAGGAAACCTTGTATGTGTGGAAT 2700  
QY 2701 GAACCTAAATGGTGCATTAAGGAATTTCTTTGCTGAGGAGCCGGTGTGGAGCCTTTTACTGCCGTTTGTGCTG 2820  
Db 2701 GAACCTAAATGGTGCATTAAGGAATTTCTTTGCTGAGGAGCCGGTGTGGAGCCTTTTACTGCCGTTTGTGCTG 2820  
QY 2821 GTGGCTCTGACCTGCTACTTCTGGAAGAAATCAAAA----- 2858  
Db 2821 GTGGCTCTGACCTGCTACTTCTGGAAGAAATCAAAAACCTGGAATACAAATATTCCCAAG 2880  
QY 2859 ----- 2858  
Db 2881 TTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCCGCTGCAGACAGTTGTGCTATC 2940  
QY 2859 ----- 2858  
Db 2941 ATGGAAGGAGAAGATAATGAAGAGGAAGTTGTATATTTCCAATAAACAGTCACTACTAGGA 3000  
QY 2859 -----GAAAAAGAAGACCATTGTAATCTGTTCAACTGAAA 2894  
Db 3001 AAACCTCAAAATCTTTGGCAACCAGGAAAAAGAGACCATTGTAATCTGTTCAACTGAAA 3060  
QY 2895 ACCTCAAGATCCCCAAATATATGAAGAGACAGTGCTGTAGCCTTGAGACTAATGAACAAA 2954  
Db 3061 ACCTCAAGATCCCCAAATATATGAAGAGACAGTGCTGTAGCCTTGAGACTAATGAACAAA 3120  
QY 2955 GAAACCTGCTCTAGTTTTACAGGACCATATTTTAGGGTCTGCTCCTCATACCTGTGCACATT 3014  
Db 3121 GAAACCTGCTCTAGTTTTACAGGACCATATTTTAGGGTCTGCTCCTCATACCTGTGCACATT 3180  
QY 3015 GGTGATCTCACAGAGGAGGGCCCTGAAAGGGAGGAGATTGAAACATTTTGATT 3074  
Db 3181 GGTGATCTCACAGAGGAGGGCCCTGAAAGGGAGGAGATTGAAACATTTTGATT 3240  
QY 3075 GCCTTATCACATGGTCAAGTACCTTGCCCAAATAAAGGAAAGCAATGATTTGGGTCTCAA 3134  
Db 3241 GCCTTATCACATGGTCAAGTACCTTGCCCAAATAAAGGAAAGCAATGATTTGGGTCTCAA 3300  
QY 3135 CTGAAGATGAAGCTCAACTCAGGAAGAGATTTATCTGTATATACACATAAAGTGAACCA 3194  
Db 3301 CTGAAGATGAAGCTCAACTCAGGAAGAGATTTATCTGTATATACACATAAAGTGAACCA 3360  
QY 3195 AGTTTAAGCCCAACCAATGCACCTGCTGATGCATGCCATATAATTAATGGTAACTTTATT 3254  
Db 3361 AGTTTAAGCCCAACCAATGCACCTGCTGATGCATGCCATATAATTAATGGTAACTTTATT 3420  
QY 3255 CTTTATGATGTCTACATAACAAGTGTGATTGGAAGGCACATGTGAGCATATGCATTATG 3314  
Db 3421 CTTTATGATGTCTACATAACAAGTGTGATTGGAAGGCACATGTGAGCATATGCATTATG 3480  
QY 3315 ATCCAATTTATGTTTTTCTTTGTTTATATTTTGGGAAAAATTAATAATTTTAAAGGTA 3374  
Db 3481 ATCCAATTTATGTTTTTCTTTGTTTATATTTTGGGAAAAATTAATAATTTTAAAGGTA 3540  
QY 3375 AAAAAAAAAAAAAA 3390  
Db 3541 AAAAAAAAAAAAAA 3556

RESULT 3

US-10-002-050-19  
; Sequence 19, Application US/10002050  
; Publication No. US20030032095A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: No. US20030032095A1el Nucleic Acid Sequences Encoding Human Se  
; FILE REFERENCE: 15966-554 Cura-54 CON-S14  
; CURRENT APPLICATION NUMBER: US/10/002,050  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 1737  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (296)..(1687)  
; NAME/KEY: variation  
; LOCATION: (1)..(1737)  
; OTHER INFORMATION: N may be any nucleotide  
US-10-002-050-19

Query Match 44.2%; Score 1499.6; DB 9; Length 1737;  
Best Local Similarity 98.7%; Pred. No. 0;  
Matches 1545; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

QY 189 GAAAGATTATCAGTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 248  
Db 142 GAAAGATTATCAGTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 201  
QY 249 CATTCACAAATTCAGAGTGACTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGTGAGAGGCAAGAAATG 308  
Db 202 CATTCACAAATTCAGAGTGACTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGTGAGAGGCAAGAAATG 261  
QY 309 CACTTTCCTCTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGTGAGAGGCAAGAAATG 368  
Db 262 CACTTTCCTCTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGTGAGAGGCAAGAAATG 321  
QY 369 TGGTGAAGGCACCTATTTCCTTTGGCAGTGGCATCAAAATTTGATGAATGGGATGAATGGCC 428  
Db 322 TGGTGAAGGCACCTATTTCCTTTGGCAGTGGCATCAAAATTTGATGAATGGGATGAATGGCC 381  
QY 429 GGCAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGACAGCAG 488  
Db 382 GGCAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGACAGCAG 441  
QY 489 GCCAGACGGCTGTAACAACTCTTCTGGATCTGTCACCTTAAGAAGTCAAGGATTAATCG 548  
Db 442 GCCAGACGGCTGTAACAACTCTTCTGGATCTGTCACCTTAAGAAGTCAAGGATTAATCG 501  
QY 549 TGATGACTGCACGGTGTCTTTGATCTATGTCACCTTAAGAAGTCAAGGATTAATCG 608  
Db 502 TGATGACTGCACGGTGTCTTTGATCTATGTCACCTTAAGAAGTCAAGGATTAATCG 561  
QY 609 CTTTGAAGTACAGTATGTCGACACACACATCTTCTTTGAGTCTTTTATTTCAAAATGATCA 668  
Db 562 CTTTGAAGTACAGTATGTCGACACACACATCTTCTTTGAGTCTTTTATTTCAAAATGATCA 621  
QY 669 GTGCCAGGAGATGGACACCACTGACAAAGTGGTAAACCTTACAGACAATGGAGAATG 728  
Db 622 GTGCCAGGAGATGGACACCACTGACAAAGTGGTAAACCTTACAGACAATGGAGAATG 681  
QY 729 GGGCTCTCATTTCTGTAATGCTGAATCAGGCAACACATCTCTACTGGAGAACTACAGG 788

Db 682 GGGCTCTCATTTCTGTAATGCTGAATCAGGCACAAACATATCTACTTGGAGAACTACAGG 741  
QY 789 CATCCTTATGGGTTCTAAGCGGTCGAAGCCTGTGCTGGTAAAAAATATCACAATGAAGG 848  
Db 742 CATCCTTATGGGTTCTAAGCGGTCGAAGCCTGTGCTGGTAAAAAATATCACAATGAAGG 801  
QY 849 GGTGGCTACACATCAGAAATGTTTCCTTGCAAGCCAGGCACATTCAGCAACAACCCAGG 908  
Db 802 GGTGGCTACACATCAGAAATGTTTCCTTGCAAGCCAGGCACATTCAGCAACAACCCAGG 861  
QY 909 TTCATTCAACTGCCAGGTGTGCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAATG 968  
Db 862 TTCATTCAACTGCCAGGTGTGCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAATG 921  
QY 969 TATAAGGTGTAAGACGACTCTCAATTTTC-----AGGATCCAGTGAGTGTACAGAGCG 1022  
Db 922 TATAAGGTGTAAGACGACTCTCAATTTTCAGAGGAAGGATCCAGTGAGTGTACAGAGCG 981  
QY 1023 CCCTCCCTGTACCACAAAAGACATTTCCAGATCCATACCCATGTGATGAAGAAGGAAA 1082  
Db 982 CCCTCCCTGTACCACAAAAGACATTTCCAGATCCATACCCATGTGATGAAGAAGGAAA 1041  
QY 1083 GACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGAGGATCTCACAGATGC 1142  
Db 1042 GACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGAGGATCTCACAGATGC 1101  
QY 1143 TATTAGATTGCCCCCTTCTGGAGAGAAGAGGATTGTCCGCCCTTGCAACCCCTGGATTTTA 1202  
Db 1102 TATTAGATTGCCCCCTTCTGGAGAGAAGAGGATTGTCCGCCCTTGCAACCCCTGGATTTTA 1161  
QY 1203 TAACAAATGGATCATCTTCTTGCCATCCCTGTCCCTCGCTGGAACATTTTCAGATGGAACCAA 1262  
Db 1162 TAACAAATGGATCATCTTCTTGCCATCCCTGTCCCTCGCTGGAACATTTTCAGATGGAACCAA 1221  
QY 1263 AGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACATTGGCTTTGAATATAAAATGGTG 1322  
Db 1222 AGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACATTGGCTTTGAATATAAAATGGTG 1281  
QY 1323 GAATGTCTCTTCCCTGGCAACATGAAAACTTCCCTGCTTCAATGTGGGAATTCAAAAGTGGGA 1382  
Db 1282 GAATGTCTCTTCCCTGGCAACATGAAAACTTCCCTGCTTCAATGTGGGAATTCAAAAGTGGGA 1341  
QY 1383 TGGAAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGA 1442  
Db 1342 TGGAAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGA 1401  
QY 1443 CAATGATTAACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATGAC 1502  
Db 1402 CAATGATTAACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATGAC 1461  
QY 1503 TGGAGCCACGGGTTCTGAACTAGGAAGAATAACATTTTGTCTTTGAGACCCCTCTGTTTCAGC 1562  
Db 1462 TGGAGCCACGGGTTCTGAACTAGGAAGAATAACATTTTGTCTTTGAGACCCCTCTGTTTCAGC 1521  
QY 1563 TGACTGTGTTTGTACTTTCATGGTGGATATTAATAGAAAAAGTACAAATGTGGTAGAATC 1622  
Db 1522 TGACTGTGTTTGTACTTTCATGGTGGATATTAATAGAAAAAGTACAAATGTGGTAGAATC 1581  
QY 1623 GTGGGGTGGAAACCAAGAAACAAAGCTTACACCCCATATCATCTTCAAGAAATGCAACTTTT 1682  
Db 1582 GTGGGGTGGAAACCAAGAAACAAAGCTTACACCCCATATCATCTTCAAGAAATGCAACTTTT 1641  
QY 1683 TACATTTACATGGG--CATTCAGAGAACTAAAT--CAGGGTCAAGATAAATAGACGGTTCA 1738  
Db 1642 TACATTTACATGGGGCATTCGCCAGAGAACTAAATTCAGGGTCCAAAGATAAATAGACGGTTCC 1701  
QY 1739 TCAAT 1743  
Db 1702 NCCAT 1706

RESULT 4

US-10-002-304-19  
; Sequence 19, Application US/10002304  
; Publication No. US20030036185A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby.  
; FILE REFERENCE: 15966-554 Cura-54 CON-S8  
; CURRENT APPLICATION NUMBER: US/10/002,304  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 1737  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (296)..(1687)  
; NAME/KEY: variation  
; LOCATION: (1)..(1737)  
; OTHER INFORMATION: N may be any nucleotide  
US-10-002-304-19

Query Match 44.2%; Score 1499.6; DB 9; Length 1737;  
Best Local Simlarity 98.7%; Pred. No. 0;  
Matches 1545; Conservative 0; Mismatches 10; Indels 10; Gaps 3;  
QY 189 GAAAGATTATCACTTTGAAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 248  
Db 142 GAAAGATTATCACTTTGAAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 201  
QY 249 CATTTCCAAATTTCTGCAGTGGACTGCTCTGGCCTGCCTGACCCAGTGAGAGGCAAGAATG 308  
Db 202 CATTTCCAAATTTCTGCAGTGGACTGCTCTGGCCTGCCTGACCCAGTGAGAGGCAAGAATG 261  
QY 309 CACTTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGCAGTAAGTG 368  
Db 262 CACTTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGCAGTAAGTG 321  
QY 369 TGGTGAAGGCACCTATTCCCTTGGCGAGTGGCATCAATTTGATGAATGGGATGAATGGCC 428  
Db 322 TGGTGAAGGCACCTATTCCCTTGGCGAGTGGCATCAATTTGATGAATGGGATGAATGGCC 381  
QY 429 GGCAGATTTTCTAACAATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTTGACAGCAG 488  
Db 382 GGCAGATTTTCTAACAATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTTGACAGCAG 441  
QY 489 GCCAGACGGCTGTAACAACCTCTTCTTGGATCCCCTCGTGGAAACTACATAGAATCTAATCG 548  
Db 442 GCCAGACGGCTGTAACAACCTCTTCTTGGATCCCCTCGTGGAAACTACATAGAATCTAATCG 501  
QY 549 TGATGACTGCACGGTGTCTTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGTCIT 608  
Db 502 TGATGACTGCACGGTGTCTTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGTCIT 561  
QY 609 CTTTGTGATACCAGTATGTCGACAAACAACATCTTCTTTGAGTTCTTTTATTCAAAAATGATCA 668  
Db 562 CTTTGTGATACCAGTATGTCGACAAACAACATCTTCTTTGAGTTCTTTTATTCAAAAATGATCA 621  
QY 669 GTGCCAGGAGATGGACACCACCACCAAGTGGTAAAACTTACAGACAATGGAGAATG 728  
Db 622 GTGCCAGGAGATGGACACCACCACCAAGTGGTAAAACTTACAGACAATGGAGAATG 681  
QY 729 GGGCTCTCATTTCTGTAATGCTGAATCAGGCACAAACATCTCTACTGGAGAACTACAGG 788

Db 682 GGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATACTCTACTGGAGAACTACAGG 741  
Qy 789 CATCCTTATGGGTTCTAAGGCGGTCAAGCCCTGTGCTGGTAAATAATATCACAATTGAAGG 848  
Db 742 CATCCTTATGGGTTCTAAGGCGGTCAAGCCCTGTGCTGGTAAATAATATCACAATTGAAGG 801  
Qy 849 GGTGGCGTACACATCAGAAATGTTTCTTGAAGCCAGGCACATTCAGCAACAACCCAGG 908  
Db 802 GGTGGCGTACACATCAGAAATGTTTCTTGAAGCCAGGCACATTCAGCAACAACCCAGG 861  
Qy 909 TTCATTCAACTGCCAGGTGTGCCAGAAACACCTATTCTGAGAAAGGAGGCCAAAGAATG 968  
Db 862 TTCATTCAACTGCCAGGTGTGCCAGAAACACCTATTCTGAGAAAGGAGGCCAAAGAATG 921  
Qy 969 TATAAGGTGTAAGACGACTCTCAATTTTC-----AGGATCCAGTGAGTGTACAGAGCG 1022  
Db 922 TATAAGGTGTAAGACGACTCTCAATTTTTCAGAGGAAGGATCCAGTGAGTGTACAGAGCG 981  
Qy 1023 CCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATACTCCATGTGATGAAGAGGAAA 1082  
Db 982 CCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATACTCCATGTGATGAAGAGGAAA 1041  
Qy 1083 GACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGAGGATCTCACAGATGC 1142  
Db 1042 GACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGAGGATCTCACAGATGC 1101  
Qy 1143 TATTAGATTGCCCCCTTCTGGAGAGAAGAGGATTGTCCGCCCTTGCAACCCCTGGATTTTA 1202  
Db 1102 TATTAGATTGCCCCCTTCTGGAGAGAAGAGGATTGTCCGCCCTTGCAACCCCTGGATTTTA 1161  
Qy 1203 TAACAATGGATCATCTTCTTGCCATCCCTGTCTCTGGAACATTTTCAGATGGAACCAA 1262  
Db 1162 TAACAATGGATCATCTTCTTGCCATCCCTGTCTCTGGAACATTTTCAGATGGAACCAA 1221  
Qy 1263 AGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGCTTGAATATAAATGGTG 1322  
Db 1222 AGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGCTTGAATATAAATGGTG 1281  
Qy 1323 GAATGTCTCTCTGGCAACATGAAAACCTTCTGTCTTCAATGTTGGGAATTCAAAGTGCGA 1382  
Db 1282 GAATGTCTCTCTGGCAACATGAAAACCTTCTGTCTTCAATGTTGGGAATTCAAAGTGCGA 1341  
Qy 1383 TGGAAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGA 1442  
Db 1342 TGGAAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGA 1401  
Qy 1443 CAATGATTACCTGATCTTAACTTGCATATCCCAGGATTTAAACCCACCAACATCTATGAC 1502  
Db 1402 CAATGATTACCTGATCTTAACTTGCATATCCCAGGATTTAAACCCACCAACATCTATGAC 1461  
Qy 1503 TGGAGCCACGGGTTCTGAACCTAGGAAGAATAACATTTGCTTTGAGACCCCTCTGTTTCAGC 1562  
Db 1462 TGGAGCCACGGGTTCTGAACCTAGGAAGAATAACATTTGCTTTGAGACCCCTCTGTTTCAGC 1521  
Qy 1563 TGACTGTGTTTGTACTTTCATGFTGGATATTAATAGAAAAAGTACAAAATGTGGTAGAATC 1622  
Db 1522 TGACTGTGTTTGTACTTTCATGFTGGATATTAATAGAAAAAGTACAAAATGTGGTAGAATC 1581  
Qy 1623 GTGGGTGGAAACCAAGAAAACAAAGCTTACACCCATATCATCTTCAAGAAATGCAACTTT 1682  
Db 1582 GTGGGTGGAAACCAAGAAAACAAAGCTTACACCCATATCATCTTCAAGAAATGCAACTTT 1641  
Qy 1683 TACATTTACATGGG--CATTTCCAGAACTAAT--CAGGGTCAAGATAATAGACGGTTCA 1738  
Db 1642 TACATTTACATGGGGCATTTCCCGAGAACTAATTCAGGGTCCCAAGATAATAGACGGTTCC 1701  
Qy 1739 TCAAT 1743  
Db 1702 NCCAT 1706

RESULT 5

US-10-003-152-19  
; Sequence 19, Application US/10003152  
; Patent No. US20020151494A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: No. US20020151494A1el Amino Acid Sequences for Human Semaphori  
; FILE REFERENCE: 15966-554 Cura-54 CON-S12  
; CURRENT APPLICATION NUMBER: US/10/003,152  
; CURRENT FILING DATE: 2001-11-02-  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 1737  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (296)..(1687)  
; NAME/KEY: variation  
; LOCATION: (1)..(1737)  
; OTHER INFORMATION: N may be any nucleotide  
US-10-003-152-19

Query Match 44.2%; Score 1499.6; DB 12; Length 1737;  
Best Local Similarity 98.7%; Pred. No. 0;  
Matches 1545; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

Qy 189 GAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGTGGAGAGTTGC 248  
Db 142 GAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGTGGAGAGTTGC 201  
Qy 249 CATTCCAAATTTCTGCAGTGGACTGCTCTGCCTGCCTGACCCAGTGAGAGGCAAGAATG 308  
Db 202 CATTCCAAATTTCTGCAGTGGACTGCTCTGCCTGCCTGACCCAGTGAGAGGCAAGAATG 261  
Qy 309 CACTTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCAGGTATGCAGTAAGTG 368  
Db 262 CACTTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCAGGTATGCAGTAAGTG 321  
Qy 369 TGGTGAGGCACCTATTCTTGGCAGTGGCATCAAATTTGATGAATGGGATGAATTGCC 428  
Db 322 TGGTGAGGCACCTATTCTTGGCAGTGGCATCAAATTTGATGAATGGGATGAATTGCC 381  
Qy 429 GGCAGGATTTCTAACATCGCAACATTCATGGACACTGTGGTGGCCCTTCTGACAGCAG 488  
Db 382 GGCAGGATTTCTAACATCGCAACATTCATGGACACTGTGGTGGCCCTTCTGACAGCAG 441  
Qy 489 GCCAGACGGCTGTAACAACTCTTCTTGGATCCTCTCGTGGAAACTACATAGAATCTAATCG 548  
Db 442 GCCAGACGGCTGTAACAACTCTTCTTGGATCCTCTCGTGGAAACTACATAGAATCTAATCG 501  
Qy 549 TGATGACTGCACGGTGTCTTTTGATCTATCTGTGCACCTTAAGAAGTCAGGCTATGTCTT 608  
Db 502 TGATGACTGCACGGTGTCTTTTGATCTATCTGTGCACCTTAAGAAGTCAGGCTATGTCTT 561  
Qy 609 CTTTGAGTACCAGTATGTCGACAACAACATCTTCTTTGAGTCTTTTATTTCAAAATGATCA 668  
Db 562 CTTTGAGTACCAGTATGTCGACAACAACATCTTCTTTGAGTCTTTTATTTCAAAATGATCA 621  
Qy 669 GTGCCAGGAGATGGACACCACCAGTGGTAAACTTTACAGACAATGGAGAATG 728  
Db 622 GTGCCAGGAGATGGACACCACCAGTGGTAAACTTTACAGACAATGGAGAATG 681  
Qy 729 GGGCTCTCATTCTGTAATGCTGAAATCAGGCACAAACATACTCTACTGGAGAACTACAGG 788



|||||  
Db 682 GGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATACTCTACTGGAGAACTACAGG 741  
Qy 789 CATCCTTATGGGTTCTAAGCGGTCAGCCCTGTGCTGTTAAATAATATCACAATTGAAGG 848  
Db 742 CATCCTTATGGGTTCTAAGCGGTCAGCCCTGTGCTGTTAAATAATATCACAATTGAAGG 801  
Qy 849 GGTGGCGGTACACATCAGAAATGTTTCCCTTGCAAGCCAGGCACATTCAGCAACAACCCAGG 908  
Db 802 GGTGGCGGTACACATCAGAAATGTTTCCCTTGCAAGCCAGGCACATTCAGCAACAACCCAGG 861  
Qy 909 TTCATTCAACTGCCAGGTGTGCCAGAGAAACACCTATTCTGAGAAAGGAGCCAAAGAAATG 968  
Db 862 TTCATTCAACTGCCAGGTGTGCCAGAGAAACACCTATTCTGAGAAAGGAGCCAAAGAAATG 921  
Qy 969 TATAAGGTGTAAAGACGACTCTCAATTTTC-----AGGATCCAGTGAGTGTACAGAGCG 1022  
Db 922 TATAAGGTGTAAAGACGACTCTCAATTTTCAGAGGAAGGATCCAGTGAGTGTACAGAGCG 981  
Qy 1023 CCCTCCCTGTACCACAAAGACTATTTCCAGATCCATACTCCATGTGATGAAGAGGAAA 1082  
Db 982 CCCTCCCTGTACCACAAAGACTATTTCCAGATCCATACTCCATGTGATGAAGAGGAAA 1041  
Qy 1083 GACACAGATAATGTACAAGTGGATAGAGCCCCAAAATCTGCCGGGAGGATCTCACAGATGC 1142  
Db 1042 GACACAGATAATGTACAAGTGGATAGAGCCCCAAAATCTGCCGGGAGGATCTCACAGATGC 1101  
Qy 1143 TATTAGATTGCCCCCTTCTGGAGAGAAGAGGATTGTCCGCCCTTGCAACCCCTGGATTTTA 1202  
Db 1102 TATTAGATTGCCCCCTTCTGGAGAGAAGAGGATTGTCCGCCCTTGCAACCCCTGGATTTTA 1161  
Qy 1203 TAACAAATGGATCATCTTCTTGCCATCCCTGTCCCTCGGAAACATTTTCAGATGGAACCAA 1262  
Db 1162 TAACAAATGGATCATCTTCTTGCCATCCCTGTCCCTCGGAAACATTTTCAGATGGAACCAA 1221  
Qy 1263 AGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAATGGTG 1322  
Db 1222 AGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAATGGTG 1281  
Qy 1323 GAATGTCTCTCTGGCAACATGAAAACTTCTGCTTCAATGTGGGAATTCAAAGTGCGA 1382  
Db 1282 GAATGTCTCTCTGGCAACATGAAAACTTCTGCTTCAATGTGGGAATTCAAAGTGCGA 1341  
Qy 1383 TGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGA 1442  
Db 1342 TGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGA 1401  
Qy 1443 CAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAACACCACCAACATCTATGAC 1502  
Db 1402 CAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAACACCACCAACATCTATGAC 1461  
Qy 1503 TGGAGCCACGGGTTCTGAACTAGGAAGATAACATTTGTCTTTGAGACCCCTCTGTTTCAGC 1562  
Db 1462 TGGAGCCACGGGTTCTGAACTAGGAAGATAACATTTGTCTTTGAGACCCCTCTGTTTCAGC 1521  
Qy 1563 TGACTGTGTTTGTACTTTCATGGTGGATATTAATAGAAAAAGTACAAATGTGGTAGAATC 1622  
Db 1522 TGACTGTGTTTGTACTTTCATGGTGGATATTAATAGAAAAAGTACAAATGTGGTAGAATC 1581  
Qy 1623 GTGGGTTGGAACCAAGAAAAACAAGCTTACACCCATATCATCTTTCAAGAAATGCAACTTT 1682  
Db 1582 GTGGGTTGGAACCAAGAAAAACAAGCTTACACCCATATCATCTTCAAGAAATGCAACTTT 1641  
Qy 1683 TACATTTACATGGG--CATTCCAGAGAACTAAAT--CAGGGTCAAGATAAATAGACGGTTCA 1738  
Db 1642 TACATTTACATGGGCAATTTCCAGAGAACTAAATTCAGGGTCCCAAGATAAATAGACGGTTCC 1701  
Qy 1739 TCAAT 1743  
Db 1702 NCCAT 1706

RESULT 6

US-10-002-050-9  
; Sequence 9, Application US/10002050  
; Publication No. US20030032095A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: No. US20030032095A1el Nucleic Acid Sequences Encoding Human Se  
; FILE REFERENCE: 15966-554 Cura-54 CON-S14  
; CURRENT APPLICATION NUMBER: US/10/002,050  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 1508  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (226)..(1458)  
; NAME/KEY: variation  
; LOCATION: (1)..(1508)  
; OTHER INFORMATION: N may be any nucleotide  
US-10-002-050-9

Query Match 34.9%; Score 1182.6; DB 9; Length 1508;  
Best Local Similarity 98.4%; Pred. No. 0;  
Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

Qy 506 ACTCTTCTTGGATCCCTCGTGGAACTACATAGAAATCTAATCGTGATGACTGCACGGTGT 565  
Db 230 ACTCTTCTTGGATCCCTCGTGGAACTACATAGAAATCTAATCGTGATGACTGCACGGTGT 289  
Qy 566 CTTTGTATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGTCTTCTTGTAGTACAGTATG 625  
Db 290 CTTTGTATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGTCTTCTTGTAGTACAGTATG 349  
Qy 626 TCGACAACAACATCTTCTTTGAGTTCTTTATTCAAATGATCAGTCCAGGAGATGGACA 685  
Db 350 TCGACAACAACATCTTCTTTGAGTTCTTTATTCAAATGATCAGTCCAGGAGATGGACA 409  
Qy 686 CCACCACCTGACAAGTGGGTAAAACCTTACAGACAATGGAGAAATGGGGCTCTCATTTCTGTAA 745  
Db 410 CCACCACCTGACAAGTGGGTAAAACCTTACAGACAATGGAGAAATGGGGCTCTCATTTCTGTAA 469  
Qy 746 TGCTGAAATCAGGCACAAACATCTCTACTGGAGAACTACAGGCAATCGCTTATGGGTTCTA 805  
Db 470 TGCTGAAATCAGGCACAAACATCTCTACTGGAGAACTACAGGCAATCGCTTATGGGTTCTA 529  
Qy 806 AGGCGGTCAAGCCTGTGCTGGTAAATAATATCACAATTTGAAGGGTGGCGTACACATCAG 865  
Db 530 AGGCGGTCAAGCCTGTGCTGGTAAATAATATCACAATTTGAAGGGTGGCGTACACATCAG 589  
Qy 866 AATGTTTTCTTGCAGGCCAGGCACATTTTCAGCAACAACCCAGGTTTCATTTCAACTGCCAGG 925  
Db 590 AATGTTTTCTTGCAGGCCAGGCACATTTTCAGCAACAACCCAGGTTTCATTTCAACTGCCAGG 649  
Qy 926 TGTGTCCCAGAAAACACCTATTCTGAGAAAGGAGCCAAAGAAATGTATAAGGTGTAAAGACG 985  
Db 650 TGTGTCCCAGAAAACACCTATTCTGAGAAAGGAGCCAAAGAAATGTATAAGGTGTAAAGACG 709  
Qy 986 ACTCTCAATTTTC-----AGGATCCAGTGAGTGTACAGAGCGGCCCTCCCTGTACCACAA 1039  
Db 710 ACTCTCAATTTTCAGAGGAAGGATCCAGTGAGTGTACAGAGCGGCCCTCCCTGTACCACAA 769  
Qy 1040 AAGACTATTTCAGATCCATCTCCATGTGATGAAGAGGAAAGACACAGATAATGTACA 1099

```
Db 770 AAGACTATTTCCAGATCCATACTCCATGTGATGAAGAAAGGAAAGACACAGATAATGTACA 829
QY 1100 AGTGGATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGATGCTATTAGATTGCCCCCTT 1159
Db 830 AGTGGATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGATGCTATTAGATTGCCCCCTT 889
QY 1160 CTGGAGAGAAGAGGATTGTCCGCCTTGCAACCCCTGGATTTTATAACAATGGATCATCTT 1219
Db 890 CTGGAGAGAAGAGGATTGTCCGCCTTGCAACCCCTGGATTTTATAACAATGGATCATCTT 949
QY 1220 CTTGCCATCCCTGTCTCCTGGAACATTTTCAGATGGAACCAAGAAAGATGTAGACCATGTC 1279
Db 950 CTTGCCATCCCTGTCTCCTGGAACATTTTCAGATGGAACCAAGAAAGATGTAGACCATGTC 1009
QY 1280 CAGCAGGAACGGAGCCCTGCACCTTGGCTTTGAATATAAATGGTGGATGAATGGTTGG 1339
Db 1010 CAGCAGGAACGGAGCCCTGCACCTTGGCTTTGAATATAAATGGTGGATGAATGGTTGG 1069
QY 1340 ACATGAAACTTCCCTGCTTCAATGTTGGGAATTCAAAGTGCATGGAATGAATGGTTGG 1399
Db 1070 ACATGAAACTTCCCTGCTTCAATGTTGGGAATTCAAAGTGCATGGAATGAATGGTTGG 1129
QY 1400 AGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGTTCTGACAAATGATTACCTGATCT 1459
Db 1130 AGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGTTCTGACAAATGATTACCTGATCT 1189
QY 1460 TAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGGTTCTG 1519
Db 1190 TAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGGTTCTG 1249
QY 1520 AACTAGGAAGAATAACATTTGCTCTTTGAGACCCTCTGTTTCAGCTGACTGTGTTTGTACT 1579
Db 1250 AACTAGGAAGAATAACATTTGCTCTTTGAGACCCTCTGTTTCAGCTGACTGTGTTTGTACT 1309
QY 1580 TCATGGTGGATATTAATAGAAAAAGTACAAATGTGGTAGAATCGTGGGTGGAACCAAAG 1639
Db 1310 TCATGGTGGATATTAATAGAAAAAGTACAAATGTGGTAGAATCGTGGGTGGAACCAAAG 1369
QY 1640 AAAACAAGCTTACACCCATATCACTTCAAGAAATGCAACTTTTACATTTACATGGG--C 1697
Db 1370 AAAACAAGCTTACACCCATATCACTTCAAGAAATGCAACTTTTACATTTACATGGGSCA 1429
QY 1698 ATTCCAGAAACTAAT--CAGGGTCAAGATAATAGACGGTTTCATCAAT 1743
Db 1430 TTCCCAGAACTAATTCAGGGTCCAAGATAATAGACGGTTCCNCCAT 1477
```

RESULT 7

US-10-002-304-9

; Sequence 9, Application US/10002304

; Publication No. US20030036185A1

; GENERAL INFORMATION:

; APPLICANT: Shimkets, Richard

; APPLICANT: Fernandes, Elma

; APPLICANT: Vernet, Corine

; APPLICANT: Yang, Meijia

; APPLICANT: Boldog, Ferenc

; APPLICANT: Herrmann, John

; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby

; FILE REFERENCE: 15966-554 Cura-54 CON-S8

; CURRENT APPLICATION NUMBER: US/10/002,304

; CURRENT FILING DATE: 2001-11-02

; PRIOR APPLICATION NUMBER: 09/604,286

; PRIOR FILING DATE: 2000-06-22

; PRIOR APPLICATION NUMBER: 60/140,584

; PRIOR FILING DATE: 1999-06-23

; NUMBER OF SEQ ID NOS: 49

; SOFTWARE: PatentIn ver. 2.0

; SEQ ID NO 9

; LENGTH: 1508

; TYPE: DNA

; ORGANISM: Homo sapiens

```
;
;
; NAME/KEY: CDS
; LOCATION: (226)..(1458)
; NAME/KEY: variation
; LOCATION: (1)..(1508)
; OTHER INFORMATION: N may be any nucleotide
US-10-002-304-9
```

Query Match 34.9%; Score 1182.6; DB 9; Length 1508;  
Best Local Similarity 98.4%; -Pred. No. 0;  
Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

```
QY 506 ACTCTTCTTGGATCCCTCGTGGAAACTACATAGAAATCTAATCGTGATGACTGCACGGTGT 565
Db 230 ACTCTTCTTGGATCCCTCGTGGAAACTACATAGAAATCTAATCGTGATGACTGCACGGTGT 289
QY 566 CTTTGATCTATGCTGTGCACCTTAAGAACTCAGGCTATGTCTTCTTTGAGTACCAGTATG 625
Db 290 CTTTGATCTATGCTGTGCACCTTAAGAACTCAGGCTATGTCTTCTTTGAGTACCAGTATG 349
QY 626 TCGACAACAACATCTTCTTTGAGTCTTTATTCAAAATGATCAGTGCCAGGAGATGGACA 685
Db 350 TCGACAACAACATCTTCTTTGAGTCTTTATTCAAAATGATCAGTGCCAGGAGATGGACA 409
QY 686 CCACCACCTGACAAGTGGTAAAACCTTACAGACAATGGAGAAATGGGGCTCTCTATTCTGTAA 745
Db 410 CCACCACCTGACAAGTGGTAAAACCTTACAGACAATGGAGAAATGGGGCTCTCTATTCTGTAA 469
QY 746 TGCTGAATTCAGGCACAAACATACTCTACTGGAGAACTACAGGCATCCTTATGGGTTCTA 805
Db 470 TGCTGAATTCAGGCACAAACATACTCTACTGGAGAACTACAGGCATCCTTATGGGTTCTA 529
QY 806 AGCGGTCAGCCCTGTGCTGGTAAAAAATATCACAATTTGAAGGGTGGCGTACACATCAG 865
Db 530 AGCGGTCAGCCCTGTGCTGGTAAAAAATATCACAATTTGAAGGGTGGCGTACACATCAG 589
QY 866 AATGTTTTCTTGCAGGCCAGGCACATTCAGCAACAAACCAGGTTTCATTCAACTGCCAGG 925
Db 590 AATGTTTTCTTGCAGGCCAGGCACATTCAGCAACAAACCAGGTTTCATTCAACTGCCAGG 649
QY 926 TGTGTCCAGAAAACACCTATTCAGAAAAGGAGCCAAAGAAATGTATAAGGTGTAAAGACG 985
Db 650 TGTGTCCAGAAAACACCTATTCAGAAAAGGAGCCAAAGAAATGTATAAGGTGTAAAGACG 709
QY 986 ACTCTCAATTTTC-----AGGATCCAGTGAGTGACAGAGCGCCCTCCCTGTACCACAA 1039
Db 710 ACTCTCAATTTTCAGAGGAAGGATCCAGTGAGTGACAGAGCGCCCTCCCTGTACCACAA 769
QY 1040 AAGACTATTTCCAGATCCATATCTCCATGTGATGAAGAAGAAAGACACAGATAATGTACA 1099
Db 770 AAGACTATTTCCAGATCCATATCTCCATGTGATGAAGAAGAAAGACACAGATAATGTACA 829
QY 1100 AGTGGATAGAGCCCCAAAATCTCCGGGAGGATCTCACAGATGCTATTAGATTGCCCCCTT 1159
Db 830 AGTGGATAGAGCCCCAAAATCTCCGGGAGGATCTCACAGATGCTATTAGATTGCCCCCTT 889
QY 1160 CTGGAGAGAAGAGGATTGTCCGCCTTGCAACCCCTGGATTTTATAACAATGGATCATCTT 1219
Db 890 CTGGAGAGAAGAGGATTGTCCGCCTTGCAACCCCTGGATTTTATAACAATGGATCATCTT 949
QY 1220 CTTGCCATCCCTGTCTCCTCGGAACATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1279
Db 950 CTTGCCATCCCTGTCTCCTCGGAACATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1009
QY 1280 CAGCAGGAACGGAGCCCTGCACITGGCTTTTGAATATAAATGGTGGAAATGTCCTTCCCTGGCA 1339
Db 1010 CAGCAGGAACGGAGCCCTGCACITGGCTTTTGAATATAAATGGTGGAAATGTCCTTCCCTGGCA 1069
QY 1340 ACATGAAAACCTTCCCTGCTTCAATGTGGGAATTCAAAGTGCATGGAATGAATGGTTGG 1399
Db 1070 ACATGAAAACCTTCCCTGCTTCAATGTGGGAATTCAAAGTGCATGGAATGAATGGTTGG 1129
QY 1400 AGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGTTCTGACAAATGATTACCTGATCT 1459
```

Db 1130 AGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTCTTGACAATGATTACCTGATCT 1189  
QY 1460 TAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGGTTCTG 1519  
Db 1190 TAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGGTTCTG 1249  
QY 1520 AACTAGGAAGAATAACATTTGTCTTTTGAGACCCCTCTGTTTCAGCTGACTGTGTTTGTACT 1579  
Db 1250 AACTAGGAAGAATAACATTTGTCTTTTGAGACCCCTCTGTTTCAGCTGACTGTGTTTGTACT 1309  
QY 1580 TCATGGTGGATATTAATAGAAAAAGTACAAATGTGGTAGAATCGTGGGTGGAAACCAAG 1639  
Db 1310 TCATGGTGGATATTAATAGAAAAAGTACAAATGTGGTAGAATCGTGGGTGGAAACCAAG 1369  
QY 1640 AAAAAAAGCTTACACCCCATCATCTTCAAGAAATGCAACTTTTACATTTACATGGG--C 1697  
Db 1370 AAAAAAAGCTTACACCCCATCATCTTCAAGAAATGCAACTTTTACATTTACATGGGGCA 1429  
QY 1698 ATTCCAGAGAACTAAT--CAGGGTCAAGATAAATAGACGGTTCATCAAT 1743  
Db 1430 TTCCAGAGAACTAATTTCAGGGTCCAAGATAATAGACGGTTCNCNCCAT 1477

RESULT 8

US-10-003-152-9  
; Sequence 9, Application US/10003152  
; Patent No. US20020151494A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: No. US20020151494A1e1 Amino Acid Sequences for Human Semaphorin-1  
; FILE REFERENCE: 15966-554 Cura-54 CON-S12  
; CURRENT APPLICATION NUMBER: US/10/003,152  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 1508  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (226)..(1458)  
; NAME/KEY: variation  
; LOCATION: (1)..(1508)  
; OTHER INFORMATION: N may be any nucleotide  
US-10-003-152-9

Query Match 34.9%; Score 1182.6; DB 12; Length 1508;  
Best Local Similarity 98.4%; Pred. No. 0;  
Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

QY 506 ACTCTTCTTGGATCCCTCGTGGAAACTACATAGAATCTAATCGTGTGATGACTGCACGGTGT 565  
Db 230 ACTCTTCTTGGATCCCTCGTGGAAACTACATAGAATCTAATCGTGTGATGACTGCACGGTGT 289  
QY 566 CTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGTCTTCTTTGAGTACCAGTATG 625  
Db 290 CTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGTCTTCTTTGAGTACCAGTATG 349  
QY 626 TCGACAACAACATCTTCTTTGAGTTCTTTTATCAAAATGATCAGTGCCAGGAGATGGACA 685  
Db 350 TCGACAACAACATCTTCTTTGAGTTCTTTTATCAAAATGATCAGTGCCAGGAGATGGACA 409

QY 686 CCACCACCTGACAAGTGGGTAAAACTTTACAGACAATGGAGAAATGGGGCTCTCATCTCTGTA 745  
Db 410 CCACCACCTGACAAGTGGGTAAAACTTTACAGACAATGGAGAAATGGGGCTCTCATCTCTGTA 469  
QY 746 TGCTGAAATCAGGCGACAAACATACTCTACTGGAGAACTTACAGGCATCCCTTATGGGTTCTA 805  
Db 470 TGCTGAAATCAGGCGACAAACATACTCTACTGGAGAACTTACAGGCATCCCTTATGGGTTCTA 529  
QY 806 AGGCGGTCAAGCCTGTGCTGGTAAAAAATATCACAAATTGAAGGGGTGGCGTACACATCAG 865  
Db 530 AGGCGGTCAAGCCTGTGCTGGTAAAAAATATCACAAATTGAAGGGGTGGCGTACACATCAG 589  
QY 866 AATGTTTTCCTTGCAGGCGAGGCACATTCAGCAACAAACAGGTTTCATTTCAACTGCCAGG 925  
Db 590 AATGTTTTCCTTGCAGGCGAGGCACATTCAGCAACAAACAGGTTTCATTTCAACTGCCAGG 649  
QY 926 TGTGTCCCGAGAAACACCTATTCTTGAGAAAGGAGCCAAAGAATGTATAAGGTGTAAAGACG 985  
Db 650 TGTGTCCCGAGAAACACCTATTCTTGAGAAAGGAGCCAAAGAATGTATAAGGTGTAAAGACG 709  
QY 986 ACTCTCAATTTTC-----AGGATCCAGTGTGAGTGTACAGAGCGCCCTCCCTGTACCA 1039  
Db 710 ACTCTCAATTTTCAGAGGAAAGGATCCAGTGTGAGTGTACAGAGCGCCCTCCCTGTACCA 769  
QY 1040 AAGACTATTTCAGATCCATACTCCATGTGATGAAGAAAGAACACACAGATAATGTACA 1099  
Db 770 AAGACTATTTCAGATCCATACTCCATGTGATGAAGAAAGAAAGAACACACAGATAATGTACA 829  
QY 1100 AGTGGATAGAGCCCCAAAATCTGCCGGGAGGATCTCACAGATGCTATTAGATTGCCCCCTTT 1159  
Db 830 AGTGGATAGAGCCCCAAAATCTGCCGGGAGGATCTCACAGATGCTATTAGATTGCCCCCTTT 889  
QY 1160 CTGGAGAGAAAGAGGATGTCCGCCTTGCAACCCTGGATTATATAACAATGGATCATCTTT 1219  
Db 890 CTGGAGAGAAAGAGGATGTCCGCCTTGCAACCCTGGATTATATAACAATGGATCATCTTT 949  
QY 1220 CTTGCCATCCCTGTCCCTCGTGGAAACATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1279  
Db 950 CTTGCCATCCCTGTCCCTCGTGGAAACATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1009  
QY 1280 CAGCAGGAACGGAGCCCTGCACCTTGGCTTTGAATATAAATGGTGGAAATGTCTTCCCTGGCA 1339  
Db 1010 CAGCAGGAACGGAGCCCTGCACCTTGGCTTTGAATATAAATGGTGGAAATGTCTTCCCTGGCA 1069  
QY 1340 ACATGAAAACTTCCTGCTTCAATGTGGGAATTCAAAGTCCGATGGAATGAATGGTGGG 1399  
Db 1070 ACATGAAAACTTCCTGCTTCAATGTGGGAATTCAAAGTCCGATGGAATGAATGGTGGG 1129  
QY 1400 AGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTTGACAATGATTACCTGATCT 1459  
Db 1130 AGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTTGACAATGATTACCTGATCT 1189  
QY 1460 TAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGGTTCTG 1519  
Db 1190 TAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGGTTCTG 1249  
QY 1520 AACTAGGAAGAATAACATTTGTCTTTTGAGACCCCTCTGTTTCAGCTGACTGTGTTTGTACT 1579  
Db 1250 AACTAGGAAGAATAACATTTGTCTTTTGAGACCCCTCTGTTTCAGCTGACTGTGTTTGTACT 1309  
QY 1580 TCATGGTGGATATTAATAGAAAAAGTACAAATGTGGTAGAATCGTGGGTGGAAACCAAG 1639  
Db 1310 TCATGGTGGATATTAATAGAAAAAGTACAAATGTGGTAGAATCGTGGGTGGAAACCAAG 1369  
QY 1640 AAAAAAAGCTTACACCCCATATCATCTTCAAGAATGCAACTTTTACATTTACATGGG--C 1697  
Db 1370 AAAAAAAGCTTACACCCCATATCATCTTCAAGAATGCAACTTTTACATTTACATGGGGCA 1429  
QY 1698 ATTCCAGAGAACTAAT--CAGGGTCAAGATAAATAGACGGTTCATCAAT 1743  
Db 1430 TTCCAGAGAACTAATTTCAGGGTCCAAGATAATAGACGGTTCNCNCCAT 1477



RESULT 9  
US-10-028-072-37  
; Sequence 37, Application US/10028072  
; Publication NO. US20030004311A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang  
; TITLE OF INVENTION:  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/028,072  
; CURRENT FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/049911  
; PRIOR FILING DATE: 1997-06-18  
; PRIOR APPLICATION NUMBER: 60/056974  
; PRIOR FILING DATE: 1997-08-26  
; PRIOR APPLICATION NUMBER: 60/059113  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059115  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059117  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059122  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059184  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059352  
; PRIOR FILING DATE: 1997-09-19  
; PRIOR APPLICATION NUMBER: 60/059588  
; PRIOR FILING DATE: 1997-09-19  
; PRIOR APPLICATION NUMBER: 60/059836  
; PRIOR FILING DATE: 1997-09-24  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/062285  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/062287  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/062814  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/062816  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063045  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063082  
; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: 60/063127  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063327  
; PRIOR FILING DATE: 1997-10-27  
; PRIOR APPLICATION NUMBER: 60/063329  
; PRIOR FILING DATE: 1997-10-27  
; PRIOR APPLICATION NUMBER: 60/063550  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063561  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063704

; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/063733  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/063735  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/063738  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/063755  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/064248  
; PRIOR FILING DATE: 1997-11-03  
; PRIOR APPLICATION NUMBER: 60/064809  
; PRIOR FILING DATE: 1997-11-07  
; PRIOR APPLICATION NUMBER: 60/065186  
; PRIOR FILING DATE: 1997-11-12  
; PRIOR APPLICATION NUMBER: 60/065846  
; PRIOR FILING DATE: 1997-11-17  
; PRIOR APPLICATION NUMBER: 60/066364  
; PRIOR FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: 60/066453  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/066511  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/066770  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/069212  
; PRIOR FILING DATE: 1997-12-11  
; PRIOR APPLICATION NUMBER: 60/069278  
; PRIOR FILING DATE: 1997-12-11  
; PRIOR APPLICATION NUMBER: 60/069334  
; PRIOR FILING DATE: 1997-12-11  
; PRIOR APPLICATION NUMBER: 60/069694  
; PRIOR FILING DATE: 1997-12-16  
; PRIOR APPLICATION NUMBER: 60/072320  
; PRIOR FILING DATE: 1998-01-23  
; PRIOR APPLICATION NUMBER: 60/073612  
; PRIOR FILING DATE: 1998-02-04  
; PRIOR APPLICATION NUMBER: 60/074086  
; PRIOR FILING DATE: 1998-02-09  
; PRIOR APPLICATION NUMBER: 60/074092  
; PRIOR FILING DATE: 1998-02-09  
; PRIOR APPLICATION NUMBER: 60/077791  
; PRIOR FILING DATE: 1998-03-12  
; PRIOR APPLICATION NUMBER: 60/078910  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/079294  
; PRIOR FILING DATE: 1998-03-25  
; PRIOR APPLICATION NUMBER: 60/079663  
; PRIOR FILING DATE: 1998-02-27  
; PRIOR APPLICATION NUMBER: 60/079728  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: 60/080165  
; PRIOR FILING DATE: 1998-03-31  
; PRIOR APPLICATION NUMBER: 60/081203  
; PRIOR FILING DATE: 1998-04-09  
; PRIOR APPLICATION NUMBER: 60/081229  
; PRIOR FILING DATE: 1998-04-09  
; PRIOR APPLICATION NUMBER: 60/081695  
; PRIOR FILING DATE: 1998-04-14  
; PRIOR APPLICATION NUMBER: 60/081817  
; PRIOR FILING DATE: 1998-04-15  
; PRIOR APPLICATION NUMBER: 60/081818  
; PRIOR FILING DATE: 1998-04-15  
; PRIOR APPLICATION NUMBER: 60/082999  
; PRIOR FILING DATE: 1998-04-24  
; PRIOR APPLICATION NUMBER: 60/083322  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: 60/083545  
; PRIOR FILING DATE: 1998-04-29  
; PRIOR APPLICATION NUMBER: 60/084600  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/084627  
; PRIOR FILING DATE: 1998-05-07







Qy	231	CTCCAGGTGGAGAGTTGCCAATTC	290
Db	327	TTCCAGGTGGAGGTCGCCGTGCCG	386
Qy	291	AGTGAGAGGCAAGAAATGCACATT	350
Db	387	CGTCAAGGGCACCGAGTGCCTCT	446
Qy	351	CCAGGTATGCAGTAAGTGTGGTGA	410
Db	447	CCAGTCATGTAAGCCATGCGCTGA	506
Qy	411	TGAATGGGATGAATTCGCGGCAG	470
Db	507	TGAGTGGGATGAGCTGCCCATGG	566
Qy	471	GGGCCCCTTCTGACAGCAGGCG	530
Db	567	CAGTGTCTGAGTGCCA---CCGG	623
Qy	531	CTACATAGAATCTAATCGTGACT	590
Db	624	CTACATCGCCCTCCAACACGCA	683
Qy	591	GAAGTCAGGCTATGTCTTTTGAG	650
Db	684	GCAATCTGGCACCCGTTAACTTC	743
Qy	651	CTTTTATTCAAAATGATCAGTGC	710
Db	744	TTTCGTTTCAGAAATGACCAGTG	800
Qy	711	TACAGACAAATGGAGAATGGGG	770
Db	801	CACAGAGAA---AGGATGGGAAT	857
Qy	771	CTACTGGAGAACTACAGGCATCC	830
Db	858	CTATTGGAGAACCCACAGCCTTC	917
Qy	831	AAATATCACAATTTGAAGGGGTG	890
Db	918	AAACATTGCCATACAGGGGTGSC	977
Qy	891	ATTTCAGCAACAACCCAGGTTCA	950
Db	978	GTATGCAGACAAGCAGGGCTCC	1037
Qy	951	GAAAGGAGCCAAAGAATGTATA	1007
Db	1038	TAAAGGAGAAACTTCTTGCCACC	1097
Qy	1008	TGAGTGTACAGAGCGCCCTCCCT	1067
Db	1098	TTCCCTGTAACGTGCGCCCAAG	1157
Qy	1068	TGATGAAGAAGGAAGACACAGAT	1127
Db	1158	CGATGCCAACGGAGAGACACAAC	1217
Qy	1128	GGATCTCACAGATGCTATTAGAT	1187
Db	1218	GGACCTTGAGGGGGCAGTGAAG	1277
Qy	1188	CAACCCCTGGATTTTATAACAAT	1247
Db	1278	CAACCCAGGCTTCTTCAAAAACC	1337
Qy	1248	TTTCAGATGGAAACCAAGAAATG	1307
Db	1338	CTCCAATGG---CTCAGACTGT	1394

Qy	1308	TGAATATAAATGGTGGAAATGTCC	1367
Db	1395	TGAATACAAATGGTGGAAACACG	1454
Qy	1368	GAATTCAAAGTGGATGGAATGAAT	1427
Db	1455	CAACTTCGAGTACAAGGGCATGAC	1514
Qy	1428	GGCTGGAGGTTCTGACAATGATT	1487
Db	1515	TGCTGGAGCCTCAGACAATGACT	1574
Qy	1488	ACCAACATCTATGACTGGAGCCAC	1544
Db	1575	TCCGCAGTCGGTGATGGCAGACAC	1634
Qy	1545	TGAGACCCCTCTGTTTCAGCTGAC	1604
Db	1635	TGAGACCCCTCTGTCTGTGAAC	1694
Qy	1605	TACAAATGTGGTAGAATCGTGGG	1664
Db	1695	CAACACTCCTGTGGAGACGTGGA	1754
Qy	1665	CTTCAAGAAATGCAACTTTTACA	1724
Db	1755	TGAGGAGAACACTACCACGAGCT	1814
Qy	1725	TAATAGACGGTTTCATCAATGAC	1784
Db	1815	AAGCAGGAAGTACACCAATGAC	1874
Qy	1785	TGATGGGGTGGCGTCCCTCATG	1844
Db	1875	GAATGGCGTGGCCTCCTACTG	1934
Qy	1845	GTTGTGCCCTGCCCTCCAGGCC	1904
Db	1935	CTGCACCTCTTGTCTGCTGCT	1994
Qy	1905	TCCACCTGACACCTACCTGTCC	1964
Db	1995	CCCCCTAACACAAATCTGAAAG	2054
Qy	1965	CGGCGCTGGGAGTAAAAACAAT	2024
Db	2055	TGTTCCAGGGACCAAGAACACA	2114
Qy	2025	CCATGAAAAAGAAAAATCAGAT	2084
Db	2115	ACGCAACACTCCAAACCAGGAC	2174
Qy	2085	ATTAATGAATGGCCCCAGCTTC	2144
Db	2175	TCCTGCTGAGGGGCCAAGCTTC	2234
Qy	2145	CAGTTTATGTGGGCATGAGGGG	2204
Db	2235	CAGTCTCTGTGGAAACCAGGGT	2294
Qy	2205	TACAGTAAAAAGAAATAGTGG	2264
Db	2295	CCGGATTCTTGAGGGTGAGTC	2345
Qy	2265	ATGCCAGTCAACAATTATTCCT	2324
Db	2346	CTGCCAGGCGAGTCATCATCC	2405
Qy	2325	ATCCATCATTTCTGGCAGATAC	2384
Db	2406	GCCTGTACGCTTGCTGATCGA	2465
Qy	2385	TAATATAAAAAGAAAGATATG	2444

```

Db 2466 CACCTCCAGCTGAACCTTTCCACCTGGAGTCCTTGGGAATACCGGACGTGATCTTCTT 2525
QY 2445 TTATAAGCTTCTTACAGCAACACATCTTGTATTAAATGGCCGATCAACTGCTGTGAAAT 2504
Db 2526 TTATAGTCCAATGATGTGACCCAGTCCTGCAGTTCTGGGAGATCAACCACCATCCGCGT 2585
QY 2505 GAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCCCAGCAAGTGCCCCAGC 2564
Db 2586 CAGGTGAGTCCACAGAAAACGTGCCCTGGAAGTTTGCTGTGCCAGGAACGTGCTCAGA 2645
QY 2565 AGGTACTGTGATGGGTGTACGTCTCTATTTCCTGTGGGAGAGTGCTGAAGCTTGCCCTCT 2624
Db 2646 TGGGACTGTGATGGCTGCAACTTCCACTTCCTGTGGGAGAGCGCGGCTGCTTGCCCGCT 2705
QY 2625 GTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCTGCAAGAGAGAGGATTTTCAGGAAAC 2684
Db 2706 CTGCTCAGTGGCTGACTACCATGCTATCGTCAGCAGCTGTGTGGCTGGGATCCAGANGAC 2765
QY 2685 CTTGTATGTGTGGAATGAACCTAAATGGTGCAATTAAGGAAATTTCTTGCCTGAGAAAAA 2744
Db 2766 TACTTACGTGTGNCGAGAACCCAAAGCTATGCTCTGGTGGCATTTCTCTGCCTGAGCAGAG 2825
QY 2745 GTTGGCAACCTGTGAAACGGTTTGACTTTTGGCTGAAGGTGGAGCCGGTGTGGGAGCTTT 2804
Db 2826 AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG 2885
QY 2805 TACTGCGGTTTGTGGTGGCTCTGACCTGCTACTTCTGGAAAAAAGAATCAAAAGAAAAA 2864
Db 2886 TACTGCCATCCTGCTCACCGTCTTGACCTGCTACTTTTGGAAAAAAGAANTCAAAAACTAGA 2945
QY 2865 GAAGACCATTTTGAATCTGTTCA 2887
Db 2946 GTACAAGTACTCCAAGCTGGTGA 2968
```

RESULT 11

US-10-123-904-37

; Sequence 37, Application US/10123904  
; Publication No. US20030022328A1  
; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3330R1C54  
; CURRENT APPLICATION NUMBER: US/10/123,904  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 37

; LENGTH: 3501  
; TYPE: DNA

; ORGANISM: Homo Sapien  
; FEATURE:

; NAME/KEY: unsure  
; LOCATION: 2762..2778  
; OTHER INFORMATION: unknown base

US-10-123-904-37

Query Match 26.6%; Score 901; DB 9; Length 3501;  
Best Local Similarity 60.0%; Pred. No. 2.5e-241;  
Matches 1634; Conservative 0; Mismatches 1062; Indels 27; Gaps 7;

```

QY 171 GCTTCCTCCTTGCCAGAGAGAAAGATTATCACCTTTGAATATACGGAATGTGATAGCAGTGG 230
Db 267 GCTTCATGCCTGCAAAAGAGTCTGAGTACCACATATGAGTACACGCGCTGTGACAGCACGGG 326
QY 231 CTCCAGGTGGAGAGTTGCCATTCCAAATTTCTCCTGTGCTTTCAGTGGACTGCTCTGGCCCTGCCTGACCC 290
Db 327 TTCCAGGTGGAGGGTGGCGGTGCCGCATACCCCGGCCCTGTGACACAGCCTGTCTGACCC 386
QY 291 AGTGACAGGCAAAAGAAATGCACCTTCTCCTGTGCTTCTGGAGAGATATCTAGAAATGAAGAA 350
Db 387 CGTCAAGGGCACCGAGTGTCTCTCTGCAACGCCGGGAGTTTCTGGATATGAAGGA 446
QY 351 CCAGGTATGCAGTAAAGTGTGGTGAAGGCACCTATTCTTGGGAGTGGCATCAAAATTTGA 410
Db 447 CCAGTCATGTAAGCCATGCGGTGAGGGCCGCTACTCCCTCGGCACAGGCATTCGGTTTGA 506
QY 411 TGAATGGGATGAATTGCCGGCAGGATTTCTAAACATCGCAACATTCATGGACACTGTGGT 470
Db 507 TGAGTGGGATGAGTGGCCCATGGCTTTCAGCCCTCTCAGCCACATGGAGCTGGATGA 566
QY 471 GGGCCCTTCTGACAGCAGGCCAGCGGCTGTAACAACCTCTTCTTGGATCCCTCGTGGAAA 530
Db 567 CAGTGTCTGTAGTCCA--CCGGGAAGTGTACTTCTGTCCAAGTGGGTCCCCGGGGCGGA 623
QY 531 CTACATAGAATCTAATCGTGTGATGACTGACACGGTGTCTTTTGATCTATGCTGTGCACCTTAA 590
Db 624 CTACATCGCCTCCAACACGGACGGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 683
QY 591 GAAGTCAGGCTATGTCTTTTGTAGTACCAGTATGTGACAAACAAACATCTTCTTTGAGTT 650
Db 684 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 743
QY 651 CTTTATTCAAAATGATCAGTGCCAGGAGATGGACACCACACTGACAAAGTGGGTAAAACT 710
Db 744 TTTTCGTTCAGAAATGACCAGTGCCAGCCCAATGCAGATGACTC--CAGGTGGATGAAGAC 800
QY 711 TACAGACAATGGAGAATGGGGCTCTCATTTCTGTAATGCTGAATCAGGCACAAACATACT 770
Db 801 CACAGAGAA--AGGATGGGAATTCACAGTGTGGAGCTAAATCGAGGCAATAATGTCT 857
QY 771 CTACTGGAGAACTACAGGCATCCTTATGGGTTCTAAGGGGGTCAAGCCCTGTGCTGGTAAA 830
Db 858 CTATTGGAGAACCCACAGCCTTCTCAGTATGGACCAAAAGTACCCAAAGCCTGTGCTGGTGAG 917
QY 831 AAATATCACAATTTGAAGGGTGGCTACACATCAGAAATGTTTTCTTGCAGCCAGGCAC 890
Db 918 AAACATTGCCATAAACAGGGGTGGCCTACACTTCAGAAATGCTTCCCTGCAAAACCTGGCAC 977
QY 891 ATTACAGCAACAAACCAGGTTTCATTCAACTGCCAGGTGTGTCACAGAAACACCTATTCTGA 950
Db 978 GTATGCAGACAAGCAGGGCTCCTCTTCTGCAAACTTTGCCACGCAACTCTTATTCAA 1037
QY 951 GAAAGGAGCCAAAGAAATGTATAAGGTGTA--GACGACTCTCAATTTTCAGGATCCAG 1007
Db 1038 TAAAGGAGAAACTTCTTGCCACCAGTGTGACCCCTGACAAATACTCAGAGAAAGGATCTTC 1097
QY 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATCTCCATG 1067
Db 1098 TTCCTGTAAAGTGGCGCCAGCTTGACAGACAAAGATTTATTTTACACACACACACGGCCTG 1157
QY 1068 TGATGAAGAAGGAAGACACAGATATGTACAAGTGGATAGAGCCCAAAATCTGCCGGGA 1127
Db 1158 CGATGCCAACGGAGAGACACAACACTCATGTACAAATGGGCCAAGCCGCAAAATCTGTAGCGA 1217
QY 1128 GGATCTCACAGATGCTATTAGATTGCCCTTCTGGAGAGAAGAGGATGTCCGCTTG 1187
Db 1218 GGACCTTGAGGGGGCAGTGAAGCTGCCTGCCTCTGGTGTGAAGACCCACTGCCACCCCTG 1277
```

QY	1188	CAACCCCTGGATT	TATATAACAATGGATCATCTTCTTGCCATCCCTGTCTCCTGGAAACAT	T	1247
Db	1278	CAACCCAGGCT	TCTTCAAAACCAACAACAGCACCTGCCAGCCCTGCCCATATGTTCCCTA	T	1337
QY	1248	TTCAGATGGAACCA	AAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCCTGCAC	T	1307
Db	1338	CTCCAATGG	--CTCAGACTGTACCCGCTCCCTGCAGGACTGAACCTGCTGTGGGAT	T	1394
QY	1308	TGAATATAAAT	GGTGGAAATGTCCCTTCCCTGGCAACATGAAAACTTCCCTGCCTTCAATGTTGG		1367
Db	1395	TGAATACAAAT	GGTGGAAACACGCTGCCCAACAACATGGAAACGACCGTTCTCAGTGGGAT		1454
QY	1368	GAATTCAAAGT	GCGATGGAATGAATGGTTGGAGGTGGCTGGAGATCATATCCAGAGTGG		1427
Db	1455	CAACTTCGAGT	ACAAGGCGATGACAGGCTGGGAGGTGGTGGTATCACATTTACACAGC		1514
QY	1428	GGCTGGAGGT	CTTGACAAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACC		1487
Db	1515	TGCTGGAGCCT	TCAGACAAATGACTTTCATGATTCTCACTCTGGTTGTGCCAGGATTTAGACC		1574
QY	1488	ACCAACATCT	ATGACTGGAGCCAC--GGGTTCTGAACTAGGAAGAATAACATTTGTCTT		1544
Db	1575	TCCGCAGT	CGTGATGGCAGACACAGAGAATAAAGAGGTGGCCAGAATCACATTTGTCTT		1634
QY	1545	TGAGACCC	TCTGTTTCAGCTGACTGTGTTTGTACTTTCATGGTGGATATTAAATAGAAAAAG		1604
Db	1635	TGAGACCC	TCTGTCTGTGAACGTGTGAGCTCTACTTTCATGGTGGTGTGAATTTCTAGGAC		1694
QY	1605	TACAAATGT	GGTAGAATCGTGGGTGGAACCAACCAAGAAAAACAAGCTTACACCCATATCAT		1664
Db	1695	CAACACT	CCCTGTGGAGACGTGGAAAGGTTCCAAAGGCCAAACAGTCTTATACCTACATCAT		1754
QY	1665	CTTCAAGAAT	GCAACTTTTACATTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGA		1724
Db	1755	TGAGGAGAA	CACACTACCACGAGCTTCACTGGGCTTCCAGAGGACCACCTTTTCATGAGGC		1814
QY	1725	TAATAGAC	CGGTTCAATGACATGGTGAAGATTTATTCTATCACAGCCACTAATGCAGT		1784
Db	1815	AAGCAGGA	AGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCCACCAATGTTAT		1874
QY	1785	TGATGGGT	GGGTGCTCATGCCGTGCCGTGTCCTGCTGAAACAGTCCGGTTCATC		1844
Db	1875	GAATGGCG	TGGCCTCTACTGCCGTCCCTGTGCCCTAGAAGCCTCTGATGTGGGCTCCTC		1934
QY	1845	GTGTGTC	CCCTGCCCTCCAGGCCACTACATTTGAGAAAGAAACCAACCAGTGCAAGGAATG		1904
Db	1935	CTGCACCT	CTTGTCTGCTGCTGTTACTATATTGACCGAGATTCAGGAACCTGCCACTCCTG		1994
QY	1905	TCCACCT	GCACACCTACCTGTCCATACATCAGTCTATGGCAAGAGAGGCTTGTATTCCATG		1964
Db	1995	CCCCCT	TAACACAAATCTGAAAGCCCACACGCTTATGGTGTCCAGGCCCTGTGTGCCCTG		2054
QY	1965	CGGGCCT	GGGAGTAAACAAATCAGGACCATTTCGGTTTGCTATAGTGACTGCTTTTCTA		2024
Db	2055	TGGTCC	AGGGACCAAGAACAAAGATCCACTCTCTGTGCTACAATGATTGCACCTTCTC		2114
QY	2025	CCATGAAA	AGAAATCAGATTTTGCACATATGACTTTAGCAACCTCAGCAGTGRGGCTC		2084
Db	2115	ACGCAAC	ACTCCAACAGGACTTTTCAACTACAACTTCTCCGCTTTGGCAAAACACCGTCAC		2174
QY	2085	ATTAAT	GAATGGCCCCAGCTTCACCTCCAAAGGAACAAAAATACTTCCATTTCTTCAATAT		2144
Db	2175	TCTT	GCTGGAGGCCAAGCTTCACCTCCAAGSGTTGAAATACTTCCATCACTTTACCCT		2234
QY	2145	CAGTTT	ATGTGSSCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATAAACAGACTT		2204
Db	2235	CAGTCT	CTGTGGAACACCAGGGTAGGAAATGTCTGTGTGCACCCGACAATGTCAC		2294
QY	2205	TACAGT	AAAAGAAATAGTGGCAGGTTCAGATGATTACACAAATTTGGTAGGGCATTTGT		2264
Db	2295	CCGGAT	TCCGTAGGGTGAGTCAGGG-----TTCTCCAAATCTATCACAGCCTACGT		2345

QY	2265	ATGCCAGTCAACAAATATTTCCCTTCTGAAAGTAAGGGTTTCCGAGCAGCCCTATCATCATCACA	2322
Db	2346	CTGCCAGGCAGTCATCATCCCCCAGAGGTGACAGGGCTACAAGGCCGGGGTTTCTCTCACA	2405
QY	2325	ATCCATCATTTCTGGCAGATACATTTTCATAGGAGTCAACAGTTGAAACCACATTTGAANAATAAT	2384
Db	2406	GCCGTGTGAGCCTTGCCTGATCGACTTATTGGGGTGACAACACAGATATGACTCTGGATGGAAAT	2465
QY	2385	TAATATAAAGAAGATATATTTCCAGTTTCCAACAAGCCAAATACCAGATGTGCATTTCTTT	2444
Db	2466	CACCTCCCCAGCTGAACCTTTTCCACCTGGAGTCCCTTGGGAATACCGGACGTGATCTTCTTT	2525
QY	2445	TTATAAGTCTTTACAGCAACAACATCTTCTATTAATTAATGGCCGATCAACTGCTGTGANAAT	2504
Db	2526	TTATAGGTCCAATGATGTGACCCAGTCCCTGCAGTTCTGGGAGATCAACCCACCATCCGCGT	2585
QY	2505	GAGGTGTAATCCTACTAATCTGGAGCAGGAGTGTATTTTCAGTCCCCCAGCAAGTGCCCAGC	2564
Db	2586	CAGGTGCAGTCCACAGAAAACTGTCCCCTGGAAGTTTGTCTGCCAGGAAGTGTCTCAGA	2645
QY	2565	AGGTACCTGTGATGGGTGTACGTTTCTATTTCCCTGTGGGAGAGTGTGAAGCTTGCCCTCTT	2624
Db	2646	TGGGACCTGTGATGGCTGCAACTTCCACTTCTCTGTGGGAGAGCGCGGCTGCTTGCCCGCT	2705
QY	2625	GTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCTGCAAGAGAGAGGATTTTCAGGAAC	2684
Db	2706	CTGCTCAGTGGCTGACTACCATGCTATCTGTACGACAGCTGTGTGGCTGGGATCCAGANGAC	2765
QY	2685	CTTGTATGTGTGAATGAACCTAAATGGTGCATTAAGGAATTTCTTTTGCCTGAGAAAAA	2744
Db	2766	TACTTTACGTGTGNCAGAACCCCAAGCTATGCTCTGGTGGCAATTTCTCTGCCCTGAGCAGAG	2825
QY	2745	GTTTGGCAACCTGTGAACGGTTGACTTTTGGCTGAAGGTGGGAGCGGGTGTGGGAGCTTTT	2804
Db	2826	AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAANAAGTGGGCATCTCTGCAGGCACTTG	2885
QY	2805	TACTGCCGTTTGTCTGGTGGCTCTTGACCTTGCTGACCTGCTACTTTCTGGAAAAAGAAATCAAAAGAAAA	2864
Db	2886	TACTGCCATCTGCTCACCGTCTTTGACCTGCTACTTTTGGAAAAAGAAATCAAAANACTAGA	2945
QY	2865	GAAGACCATTTTGAATCTGTTC	2887
Db	2946	GTACAAGTACTCCCAAGCTGGTGA	2968

RESULT 12

US-10-140-470-37

Sequence 37, Application US/10140470  
Publication NO. US20030022331A1

; GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

APPLICANT: Sherwood, Steven

APPLICANT: Smith, Victoria

APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

APPLICANT: Watanabe, Colin K

APPLICANT: Wood, William

APPLICANT: Zhang, Zemin

1. TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME

FILE REFERENCE: D2230D1Q160

FILE REFERENCE: P3330R1C160

;; CURRENT APPLICATION NUMBER: US/10/140,470  
;; CURRENT FILING DATE: 2002-05-05

; CURRENT FILING DATE: 2002-05-06  
 ; Prior Application removed - See

;; Prior Application removed - See palm or File Wrapper



; NUMBER OF SEQ ID NOS: 550									
; SEQ ID NO 37									
; LENGTH: 3501									
; TYPE: DNA									
; ORGANISM: Homo Sapien									
; FEATURE:									
; NAME/KEY: unsure									
; LOCATION: 2762, 2778									
; OTHER INFORMATION: unknown base									
US-10-140-470-37									
Query Match 26.6%; Score 901; DB 9; Length 3501;									
Best Local Similarity 60.0%; Pred. No. 2.5e-241;									
Matches 1634; Conservative 0; Mismatches 1062; Indels 27; Gaps 7;									
QY	171	GCTTCCTCCTGCCAGGAGAAAGATTATCACTTTGAATATACGGGAATGTGATAGCAGTGG	230						
Db	267	GCTTCATGCCCTGCAAGAGTCTGAGTACCACATATGAGTACACGGCGTGTGACAGCACGGG	326						
QY	231	CTCCAGGTGGAGAGTTGCCATTCCAAATTCTGCAGTGGACTGCTCTGGCCTGCCTGACCC	290						
Db	327	TTCCAGGTGGAGGGTCGCCGTGCCGCATACCCCGGCCCTGTGCACCCAGCCTGTCTGACCC	386						
QY	291	AGTGAGAGGCAAGAAATGCACCTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAA	350						
Db	387	CGTCAAGGGCACCGAGTGTCTCTCTGCAACGCCGGGAGTTTCTGGATATGAAGGA	446						
QY	351	CCAGGTATGCAGTAAGTGTGGTGAAGGCACCTATTCTTAAACATCGCAACATTCATGGACACTGTGGT	470						
Db	507	TGAGTGGGATGAGCTGCCCATGGCTTGGCAGCCTCTCAGCCCAACATGGAGCTGGATGA	566						
QY	471	GGGCCCTTCTGACAGCAGGCCAGACGGCTGTAAACAACCTCTTCTTGGATCCCTCGTGGAAA	530						
Db	567	CAGTGTCTGAGTCCA--CCGGGAAGTGTACTTCGTCCAAGTGGGTTCCTCCCGGGCGA	623						
QY	531	CTACATAGAAATCTAATCGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAA	590						
Db	624	CTACATCGCCTCCAACACGGACGGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA	683						
QY	591	GAAGTCAGGCTATGTCTTCTTTGAGTACCAGTATGTGCACAACAACATCTTCTTTGAGTT	650						
Db	684	GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT	743						
QY	651	CTTTATTCAAATGATCAGTGCCAGGAGATGGACACCACCTGACAAAGTGGGTAAAACT	710						
Db	744	TTTCGTTTCAGAAATGACCAAGTGCCAGGCCCAATGCAAGATGACTC---CAGGTGGATGAAGAC	800						
QY	711	TACAGACAATGGAGAATGGGGCTCTCATTTCTGTAAATGCTGAAATCAGGCACAAACATACT	770						
Db	801	CACAGAGAA--AGGATGGGAATCCACAGTGTGGAGCTAAATCGAGGCAATAATGTCT	857						
QY	771	CTACTGGAGAACTACAGGCATCCCTTATGSGTTCTAAGGCGGTCAAGCCTGTGCTGGTAAA	830						
Db	858	CTATTGGAGAACCAAGCCTTCTCAGTATGGACCAAGTACCCAAGCCTGTGCTGGTGAG	917						
QY	831	AAATATCACAAATTGAAGGGGTGGCGTACACATCAGAATGTTTTCCTTGGAAAGCCAGGCAC	890						
Db	918	AAACATTGCCATAACAGGGGTGGCCTACACTTCAGAATGCTTCCCTTGCAAAACCTGGCAC	977						
QY	891	ATTACGACAACAACCAAGGTTTCATTCAACTGCCAGGTGTGTCCCAGAAACACCTATTCTGA	950						
Db	978	GTATGCAGACAAGCAGGGCTCCTCTTTCTGCAAACTTTGCCAGCCAACTCTTATTCAA	1037						
QY	951	GAAGGAGCCAAAGAATGTATAAGGTGTA---GACGACTCTCAATTTTCAGGATCCAG	1007						
Db	1038	TAAAGGAGAAACTTCTTGCCACCAGTGTGACCTTGACAAATACTCAGAGAAAGGATCTTC	1097						
QY	1008	TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATCTCCATG	1067						

Db	1098	TTCTGTAAACGTGCGCCAGCTTGCAAGACAAAAGATTATTCTTACACACACACGCGCTG	1157						
QY	1068	TGATGAAGAAGAAAGACACAGATAATGTACAAGTGGATAGAGCCCAAAAATCTGCCGGA	1127						
Db	1158	CGATGCCAACGGAGAGACACAACACTCATATAAAATGGGCAAGCCGAAAATCTGTAGCGA	1217						
QY	1128	GGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAAGAAGGATTGTCCGCTTG	1187						
Db	1218	GGACCTTGAGGGGGCAGTGAAGCTGCCTCTGCTGTGAAGACCCACTGCCACCCCTG	1277						
QY	1188	CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTCTGGAACATT	1247						
Db	1278	CAACCCAGGCTTCTTCAAAACCAACAACAGCACCTGCCAGCCCTGCCCATATGGTTCCTA	1337						
QY	1248	TTCAGATGGAACCAAAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTT	1307						
Db	1338	CTCCAATGG--CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGATT	1394						
QY	1308	TGAATATAAATGGTGAATGTCTTCTTGGCAACATGAAAACCTTCTGCTTCAATGTTGG	1367						
Db	1395	TGAATACAAATGGTGAACACGCTGCCCAACAAACATGGAACGACCGTCTCAGTGGAT	1454						
QY	1368	GAATTCAAAGTCGATGGAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGG	1427						
Db	1455	CAACTTCGAGTACAAGGGCATGACAGGCTGGGAGGTGGCTGGTATCACATTTACACAGC	1514						
QY	1428	GGCTGGAGGTTCTGACAAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAACC	1487						
Db	1515	TGCTGGAGCCTCAGACAATGACTTCATGATCTCTACTCTGGTGTGCCAGGATTTAGACC	1574						
QY	1488	ACCAACATCTATGACTGGAGCCAC--GGGTTCTGAACCTAGGAAGAATAACATTTGTCTT	1544						
Db	1575	TCCGCAGTCGGTGATGGCAGACACAGAGAATAAAGAGGTGGCCAGAAATCACATTTGTCTT	1634						
QY	1545	TGAGACCCCTCTGTTTCACTGACTGTGTTTGTACTTTCATGGTGGATATTAATAGAAAAG	1604						
Db	1635	TGAGACCCCTCTGTTCTGTGAACCTGTGAGCTTACTTTCATGGTGGTGTGAATTTCTAGGAC	1694						
QY	1605	TACAAATGTGTFAGAATCGTGGGTGGAAACCAAGAAAACAAAGTTTACACCCATATCAT	1664						
Db	1695	CAACACTCCTGTGGAGACGTGGAAGGTTCCAAAGGCCAAACAGTCTCTATACCTACATCAT	1754						
QY	1665	CTTCAAGAAATGCAACTTTTACATTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGA	1724						
Db	1755	TGAGGAGAACACTACCACGAGCTTCACCTGSGCCTTCCAGAGGACCCTTTTCATGAGGC	1814						
QY	1725	TAATAGACGGTTTCATCAATGACATGGTGAAGATTTATTTCTATCACAGCCCACTAATGCAGT	1784						
Db	1815	AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCACCAATGTTAT	1874						
QY	1785	TGATGGGTGGGTCTCTCATGCCGTGCTGTCCTCGGTTCTTGAACAGTCGGGTTTCATC	1844						
Db	1875	GAATGGCGTGGCCTCTACTGCCGTCCCTGTGCCCTTAGAAGCCTCTGATGTGGGCTCCTC	1934						
QY	1845	GTGTGTCCCTGCCCTCCAGGCCACTACATTCAGAGTCTATGGCAAAGAGGCTTGTATTCATG	1904						
Db	1935	CTGCACCTCTGTCTCTGCTGCTGTTACTATATTGACCGAGATTCAGGAACCTGCCACTCCTG	1994						
QY	1905	TCCACCTGACACCTACCTGTCCATACATCAGTCTATGGCAAAGAGGCTTGTATTCATG	1964						
Db	1995	CCCCCTTAACACAATCTGAAAGCCCAACCGCTTATGGTGTCCAGGCCCTGTGTGCCCTG	2054						
QY	1965	CGGGCCTGGGAGTAAAAACAATCAGGACCCTTCCGGTTTGTATAGTACTGCTTTTCTA	2024						
Db	2055	TGGTCCAGGGACCAAGAACACAAGATCCACTCTCTGTCTACAATGATTCACCTTCTC	2114						
QY	2025	CCATGAAAAAGAAAATCAGATTTTGCACATATGACTTTTAGCAACCTCAGCAGTGTGGGCTC	2084						
Db	2115	ACGCAACACTCCAACCCAGGACTTTTCAACTACAACCTTCTCCGCTTTGGCAACACCCGTAC	2174						
QY	2085	ATTAATGAATGGCCCCAGCTTCACCTCCAAAGSAAACAAAATACTTCCATTTCTTCAATAT	2144						
Db	2175	TCTTGCTGGAGGGCCAAAGCTTCACTTCCAAAGGGTTGAAATACTTCCATCATCTTTACCCT	2234						



Db	978	GTATGACAGAAGCAGGGCTCCTCTTTCTGCAAACTTTGCCAGCCAACCTCTTATTCAA	1037
Qy	951	GAAGGAGCCAAAGAATGTATAAGGTGTAAA---GACGACTCTCAATTTTCAGGATCCAG	1007
Db	1038	TAAAGGAGAAACTTCTTGCCACCAGTGTGACCCTGACAAATACTCAGAGAAAGGATCTC	1097
Qy	1008	TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAAGACATATTTCCAGATCCATACCTCCATG	1067
Db	1098	TTCTCTTAACGTGCGCCACAGCTTGACACAGACAAAGATTAATTTCTACACACACACGGCCTG	1157
Qy	1068	TGATGAAGAAGGAAGACACACAGATAAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGGA	1127
Db	1158	CGATGCCAAACGGAGAGACACAACACTCATGTACAAATGGGCCAAAGCCGAAAATCTGTAGCGA	1217
Qy	1128	GGATCTCACAGATGCTATTAGATTGCCCCCTTCTTGAGAGAGAAGAGGATTGTCCGCCCTTG	1187
Db	1218	GGACCTTGAGGGGCAGTGAAGCTGCCTGCCTCTGGTGTGAAGACCCCACTGCCCAACCCTG	1277
Qy	1188	CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCCTCGGAACATT	1247
Db	1278	CAACCCAGGCTTCTTCAAAACCAACAACAGCACCTGCCAGCCCTGCCATATGGTTCCCTA	1337
Qy	1248	TTCAGATGGAACCAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCCTGCACATTGGCTT	1307
Db	1338	CTCCAAATGG---CTCAGACTGTACCCTGCCTGCCTGCAGGGACTGAACCTGCTGTGGGATT	1394
Qy	1308	TGAATATAAATGGTGGAAATGTCTCTCTGGCAACATGAAAACCTTCTGCTTCAATGTGG	1367
Db	1395	TGAATACAAATGGTGGAAACACGCTGCCCAACAACATGGAACGACCGTTCTCAGTGGGAT	1454
Qy	1368	GAATTCAAAGTGCATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGG	1427
Db	1455	CAACTTCGAGTACAAGGGCATGACAGGCTGGGAGGTGGCTGTGTGATCACATTTACACAGC	1514
Qy	1428	GGCTGGAGGTTCTGACAATGATTACCTGATCTTAAACTTGATATCCAGGATTTAAACC	1487
Db	1515	TGCTGGAGCCTCAGACAATGACTTCATGATTCTCACTCTGTTGTGCCAGGATTTAGACC	1574
Qy	1488	ACCAACATCTATGACTGGAGCCAC--GGGTTCTGAAC TAGGAATAAACATTTGTCTT	1544
Db	1575	TCCGCAGTCGGTGATGGCAGACACAGAGATAAAGAGGTGGCCAGAAATCACATTTGTCTT	1634
Qy	1545	TGAGACCCCTCTGTTCAGCTGACTGTGTTTTGTACTTTCATGTGGATATTAAATAGAAAAG	1604
Db	1635	TGAGACCCCTCTGTTCTGTGAACGTGAGCTCTACTTTCATGGTGGGTGTAATTTCTAGGAC	1694
Qy	1605	TACAAATGTGGTAGAATCGTGGGTGGAAACCAAGAAAACAAGCTTACACCCCATATCAT	1664
Db	1695	CAACACTCCTGTGGAGACGTGGAAGGTTCCAAGGGCAAAACAGTCCCTATACCTACATCAT	1754
Qy	1665	CTTCAAGAATGCAACTTTTACATTTACATGGGCATTCCAGAGAACTAATCAGGGTCAAGA	1724
Db	1755	TGAGGAGAACACTACCACGAGCTTCACCTGGCCCTTCCAGAGGACCACTTTTCATGAGGC	1814
Qy	1725	TAATAGACGGTTTCATCAATGACATGGTGAAGATTTATTCTATCACAGCCCACTAATGCAGT	1784
Db	1815	AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCACCAATGTTAT	1874
Qy	1785	TGATGGGGTGGCGTCCCTCATGCCGTGCCTGTGCCCTCGGTTCTGAACAGTCGGGTTTCATC	1844
Db	1875	GAATGGCGTGGCCTCCTACTGCCGTGCCCTGTGCCCTAGAAGCCCTCTGATGTGGGCTCCTC	1934
Qy	1845	GTGTGTCCCTGCCCTCCAGGCCACTACATATTGAGAAAGAAAACAACCAAGTGCAAGGAATG	1904
Db	1935	CTGCACCTCTTGTCTCTGTGTTACTATATATTGACCGAGATTACGAAACCTGCCACTCCTG	1994
Qy	1905	TCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTCCTATG	1964
Db	1995	CCCCCTTAACACAATTTCTGAAAGCCCAACCAAGCCTTATGGTGTCCAGGCCCTGTGTGCCCTG	2054
Qy	1965	CGGGCCTGGGAGTAAAAACAATCAGGACCATTCGGTTTGTCTATAGTGACTGCTTTTCTA	2024

Db	2055	TGFTCCAGGGACCAAGAACAACAGATCCACTCTCTGTGCTACAATGATTGCACCTTCTC	2114
Qy	2025	CCATGAAAAAGAAAATCAGATTTTGCACATATGACTTTTAGCAACCTCAGCAGTGTGGGCTC	2084
Db	2115	ACGCAACACTCCAACCAGGACTTTCAACTACAACCTTCTCCGTTTGGCAACACCCGTCA	2174
Qy	2085	ATTAAATGAATGGCCCCCAGCTTCACCTCCAAGGAACAATACTTCCATTCTTCAATAT	2144
Db	2175	TCITTGCTGGAGGCCAAGCTTCACCTCCAAGGGTTGAATACTTCCATCACCTTACCCT	2234
Qy	2145	CAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTT	2204
Db	2235	CAGTCTCTGTGGAAACCAGGGTAGGAAAATGTCTGTGCGACCGACAATGTCACTGACCT	2294
Qy	2205	TACAGTAAAAGAAATAGTGGCAGGGTCAGATGATTACACAAAATTTGGTAGGGCATTTGT	2264
Db	2295	CCGATTCCTGAGGGTGAGTCAGGG-----TTCTCCAATCTATCACAGCCTACGT	2345
Qy	2265	ATGCCAGTCAACAATTAATTCCTTCTGAAAAGTAAGGGTTTCCGAGCAGCCTTATCATCACA	2324
Db	2346	CTGCCAGGCAGTCATCATCCCCCAGAGGTCACAGGCTACAAGCGGGGTTTCTCTCACA	2405
Qy	2325	ATCCATCATTTCTGGCAGATACATTCATAGGAGTCACAGTTGAAACCACATTTGAAAAATAT	2384
Db	2406	GCCTGTCAGCCTTGTGATCGACTTATTTGGGTGACAACAGATATGACTCTGGATGGAAT	2465
Qy	2385	TAATATAAAGAAGATATGTTCCAGTTCCAACAAGCCAAATACCAGATGTGCATTTCTT	2444
Db	2466	CACCTCCCCAGCTGAACCTTTTCCACCTGGAGTCCTTTGGGAATACCGGACGTGATCTTCT	2525
Qy	2445	TTATAAGTCTTCTACAGCAACAACATCTTGTATTAAATGGCCGATCAACTGCTGTGAAAAT	2504
Db	2526	TTATAGGTCCAATGATGTGACCCAGTCCTGCAGTTCTGGGAGATCAACCACCATCCGCGT	2585
Qy	2505	GAGTGTAAATCCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCCCAGCAAGTGCCACG	2564
Db	2586	CAGTGCAGTCCACAGAAAACGTCCCTGGAAGTTTGTGCTGCCAGGAACGTGCTCAGA	2645
Qy	2565	AGTACCTGTGATGGGTGACGTTCCTATTCTATTCTGTGGGAGAGTGTGAAGCTTGCCCTCT	2624
Db	2646	TGGGACCTGTGATGGCTGCAACTTCCACTTCCACTTCTGTGGGAGAGCGGGCTGCTTGCCCGCT	2705
Qy	2625	GTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCCTGCAAGAGAGGATTTTCAGGAAAC	2684
Db	2706	CTGCTCAGTGGCTGACTACCATGCTATCGTCAGCAGCTGTGTGCTGGGATCCAGANGAC	2765
Qy	2685	CTTGTATGTGTGAATGAACCTAAATGGTGCATTAAGGAATTTCTTTGCCCTGAGAAAA	2744
Db	2766	TACTTACGTGTGNCGAGAACCCAAAGCTATGCTCTGTTGGCAATTTCTCTGCCTGAGCAGAG	2825
Qy	2745	GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGGTGGAGCCGGTGTGGGAGCTTT	2804
Db	2826	AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG	2885
Qy	2805	TACTGCCGTTTTGCTGGTGGCTCTGACCTGCTACTTCTGGAAGAAAGAAATCAAGAAAA	2864
Db	2886	TACTGCCATCCTGCTCACCCGCTTTCACCTGCTACTTTTGAAAAAGAAATCAAAAACTAGA	2945
Qy	2865	GAAGACCATTTTGAATCTGTTCA	2887
Db	2946	GTACAAGTACTCCAAGCTGGTGA	2968

RESULT 14  
US-10-176-918-37  
; Sequence 37, Application US/10176918  
; Publication NO. US20030027275A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen



```
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C382
; CURRENT APPLICATION NUMBER: US/10/176,918
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 37
; LENGTH: 3501
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 2762, 2778
; OTHER INFORMATION: unknown base
US-10-176-918-37

Query Match      26.6%; Score 901; DB 9; Length 3501;
Best Local Similarity 60.0%; Pred. No. 2.5e-241;
Matches 1634; Conservative 0; Mismatches 1062; Indels 27; Gaps 7;

Qy 171 GCTTCCTCCTGCCAGGAGAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGG 230
      ||||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 267 GCTTCATGCCTGCAAGAGTCTGAGTACCACCTATGAGTACACGGCGTGTGACAGCACGGG 326

Qy 231 CTCAGGTGGAGAGTTGCCATTCCAAATTCGACAGTGGACTGCTCTGGCCTGCCGTGACCC 290
      ||||| ||||| || ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 327 TTCCAGGTGGAGGGTCGCCGTGCCGCATACCCCGGCCCTGTGCACCAAGCCTGTCTGACCC 386

Qy 291 AGTGAGAGGCAAGAAGATGCACTTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAA 350
      || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 387 CGTCAAGGCGACCGAGTGTCTCCTCTCTCTGCAACGCGCGGGAGTTCTGGATATGAAGGA 446

Qy 351 CCAGGTATGCAGTAAGTGTGGTGAAGGCACCTATTCCCTTGGSCAGTGGCATCAAAATTGA 410
      |||| | ||| | || | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 447 CCAGTCATGTAAGCCATGCGCTGAGGCGCGCTACTCCCTCGSCACAGGCATTCCGTTTGA 506

Qy 411 TGAATGGGATGAATTGCCGGCAGGATTTTCTAACATCGCAACATTCATGGACACTGTGCT 470
      ||| ||||| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 507 TGAGTGGGATGAGCTGCCCATGGCTTTGCCAGCCTCTCAGGCAACATGGAGCTGGATGA 566

Qy 471 GGGCCCTTCTGACAGCAGGCCGAGCGGCTGTAACAACTCTTCTGGATCCCTCGTGGAAA 530
      | ||||| || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 567 CAGTGTGCTGAGTCCA ---CCGGGAAGTGTACTTTCGTCCCAAGTGGGTTCCTCCGGGGCGA 623

Qy 531 CTACATAGAATCTAATCGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAA 590
      ||||| | ||| | || | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 624 CTACATCGCTCCAACACGGACGGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 683

Qy 591 GAAGTCAGGCTATGTCTTCTTTGAGTACCAGTATGTCGACAAACAACTCTTCTTTGAGTT 650
      | ||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 684 GCAATCTGCGACCCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 743

Qy 651 CTTTATTCAAAATGATCAGTGCCAGGAGATGGACACCACCTACTGACAAGTGGGTAAAACT 710
      || |||| ||||| ||||| || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 744 TTTTCGTTCAGAATGACCAGTGCCAGGCCCAATGCAGATGACTC ---CAGGTGGATGAAGAC 800

Qy 711 TACAGACAATGGAGAATGGGGCTCTCATTTCTGTAATGCTGAATCAGGCACAAACATACT 770
      ||||| || ||||| || ||| |||| | || ||| |||| | |||| | |||| | |||| | |||| |
Db 801 CACAGAGAA ---AGGATGGGAATTCACACAGTGTGGAGCTAATCGAGGGCAATAATGTCT 857
```

```
Qy 771 CTACTTGGAGAACTACAGGCACTCCTTTATGGGTTCATAAGGGGTCAGCCCTGTGCTGGTAAA 830
      ||| ||||| |||| | || | || | ||| ||||| ||||| |
Db 858 CTATTGGAGAACCAACAGCCCTTCTCAGTATGGACCACAAAGTACCCAAGCCCTGTGCTGGTGAG 917

Qy 831 AAATATCACAAATTGAAGGGTGGCGGTACACATCAGAAATGTTTCTTGCAGGCCAGGCAC 890
      ||| | | ||||| |||| | |||| | |||| | ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 918 AAACATTGCCATAACAGGGTGGCCTACACTTCAGAAATGCTTCCCTGCMACCTGGGCAC 977

Qy 891 ATTACAGCAACAAACCAGGTTCATTTCAACTGCCAGGTGTGCCAGAGAAACACCTATTCTGA 950
      | |||| | ||| | ||| | |||| | |||| | ||| ||| ||| |||| |
Db 978 GTATGCAGACAAGCAGGGCTCCTCTTTCTGCMAACTTTGCCACGCCNACTCTTATTCAAA 1037

Qy 951 GAAAGCAGCCAAAGAAATGTATAAGGTGTAAA ---GACGACTCTCAATTTTCAGGATCCAG 1007
      ||||| | || | |||| | |||| | |||| | ||| ||| ||| |||| |
Db 1038 TAAAGGAGAAACTTCTTGCCACCAGTGTGACCCCTGACAAATACTCAGAGAAAGGATCTTC 1097

Qy 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCCATCTCCATG 1067
      | |||| | ||||| || ||| |||| | |||| | |||| | ||| ||| ||| ||| ||| ||| |||
Db 1098 TTCTCTGTAACTGCGGCCAGCTTGCACAGACAAAAGATTTATTTCTACACACACACGGCCTG 1157

Qy 1068 TGATGAAGAAGGAAGACACAGATAATGTACAAATGGATGGATAGAGCCCAAAATCTGCCGGGA 1127
      |||| | ||| ||||| | ||||| |||| | |||| | |||| | |||| | |||| | |||| |
Db 1158 CGATGCCAACGGAGAGACAAACTCATGTACAAATGGGCCAACCCGAAATCTGTAGCGGA 1217

Qy 1128 GGAATCTACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAAAGGATTTGCCGCCCTTGG 1187
      ||| || | || | ||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 1218 GGACCTTGAGGGGCGAGTGAAGCTGCCTGCCTCTGGTGTGAAGACCCACTGCCACCCCTG 1277

Qy 1188 CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTCTGGAACATTT 1247
      ||||| |||| | || | ||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 1278 CAACCCAGGCTTTCTCAAAACCACAAACAGCACCTGCCAGCCCTGCCCATATGGTTCCCTA 1337

Qy 1248 TTCAGATGGAACCAAGAAATGTAGACCATGTCCAGCAGGAAACGGAGCCCTGCACCTGGCTTT 1307
      || |||| | ||| |||| | ||| |||| | ||| |||| | ||| |||| | ||| |||| | ||| |||
Db 1338 CTCCAATGG ---CTCAGACTGTACCCGCTGCCCTGCAGGAGTGAACCTGCTGTGGGATTT 1394

Qy 1308 TGAATATAAATGGTGGAAATGCTTCTTCCCTGGCAACATGAAAATCTTCTGCTTCAATGTTGG 1367
      ||||| ||||| ||||| |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 1395 TGAATACAAATGGTGGAAACACGCTGCCCAACAACATGGAACACGACCGTTCTCAGTGGGAT 1454

Qy 1368 GAATTCAAAGTCCGATGGAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427
      || | ||| | ||| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 1455 CAACTTCGAGTACAAGGGCATGACAGGCTGGGAGGTGGCTGGTGTATCACATTTACACAGC 1514

Qy 1428 GGCTGGAGGTTCTGACAAATGATTACCTGATCTTAAACTTCGATATCCCAGGATTTTAAACC 1487
      ||||| || ||||| || |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 1515 TGCTGGAGCCTCAGACAAATGACTTCATGATTTCTCACTCTGGTGTGCCAGGATTTTAGACC 1574

Qy 1488 ACCAACAATCTATGACTGGAGCCAC ---GGGTTCGAACTAGGAAGAATAACATTTGTCTTT 1544
      || | ||| | ||| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 1575 TCCGCAGTCGGTGATGGCAGACACAGAGAATAAAGAGGTGGCCAGAAATCNCATTTGTCTTT 1634

Qy 1545 TGAGACCCCTCTGTTTCAGCTGACTGTGTTTGTACTTTCATGGTGATATTAATAGAAAAAG 1604
      ||||| ||||| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 1635 TGAGACCCCTCTGTTCTGTGAAGTGTGAGCTCTACTTTCATGGTGGGTGTGAATTTCTAGGAC 1694

Qy 1605 TACAAATGTGGTAGAATCGTGGGTGGAAACCAAGAAAGAAACAAAGCTTACACCCATATCAT 1664
      | | | ||| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 1695 CAACACTCCTGTGGAGACGTGGAAAGGTTCCAAAGGCAACAGCTCCTATACCTACATCAT 1754

Qy 1665 CTTTCAAGAATGCAACTTTTACATTTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGA 1724
      |||| | || | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 1755 TGAGGAGAACACTACCACGAGCTTCACCTGGGCCCTCCAGAGGACCACCTTTTCATGAGGC 1814

Qy 1725 TAATAGACGGTTTCATCAATGACATGGTGAAGATTTTATTTCTATCACAGGCCACTAATGCAGT 1784
      | | | ||||| || |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 1815 AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCACCAATGTTTAT 1874

Qy 1785 TGATGGGGTGGCGTCTCATGCGCTGCCCTGTGCCCTCGGTTTCTGAACAGTTCGGGTTTCAATC 1844
      |||| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 1875 GAATGGCGTGGCCTCCTACTGCCGTCCCTGTGCCCTAGAAAGCCTCTGATGTGGGCTCCTC 1934

Qy 1845 GTGTGTCCCTTGCCCTCCAGGGCCACTACATTTGAGAAAGAAACCAACCAAGTGAAGGAATG 1904
```

Db 1935 CTGCACCTCTTGTCCTGCTGGTTACTATATTGACCGAGATTTCAGGAACCTGCCACTCCTG 1994  
QY 1905 TCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGAGGCTTGATTCATG 1964  
Db 1995 CCCCCCTAACACAATCTGAAAGCCCAACAGCTTATGGTGTCCAGGCCCTGTGTGCCCTG 2054  
QY 1965 CGGGCCTGGGAGTAAAAACAATCAGGACCATTTCGGTTTGCTATAGTGACTGCTTTTCTA 2024  
Db 2055 TGGTCCAGGGACCAAGAACAACAAGATCCACTCTCTGTGCTACAATGATGCACCTTCTC 2114  
QY 2025 CCATGAAAAAGAAATCAGATTTTGGCACTATGACTTTTAGCAACCTCAGCAGTGTGGGCTC 2084  
Db 2115 ACGCAACACTCCAAACCAGGACTTTCAACTACAACTTCTCCGCTTTGGCAAAACACCGTCAC 2174  
QY 2085 ATTAATGAATGGCCCCAGCTTCACCTCCAAGGAACAAAATACTTCCATTTCTTCAATAT 2144  
Db 2175 TCTTGCTGGAGGCCAAGCTTCACCTCCAAGGGTTGAAATACTTCCATCACTTTACCCCT 2234  
QY 2145 CAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAAACAAGACTT 2204  
Db 2235 CAGTCTCTGTGAAACACAGGGTAGGAAATGTCTGTGTGCACCGACAACTGCTCACTGACCT 2294  
QY 2205 TAAGTAAAGAAATAGTGGCAGGGTCAGATGATTACACAAAATTTGGTAGGGGCATTTGT 2264  
Db 2295 CCGGATTCTGAGGGTGAGTCAGGG-----TTCTCCAAATCTATCACAGCCTACGT 2345  
QY 2265 ATGCCAGTCAACAATATTCTCTTGAAAGTAAGGGTTTCCGAGCGACCTTATCATCACACA 2324  
Db 2346 CTGCCAGGCGATCATATCCCCCAGAGGTGACAGGCTACAAGGCCGGGTTTCTCTCACA 2405  
QY 2325 ATCCATCATTTCTGGCAGATACATTTCAATAGGAGTCACAGTTTGAACCACACATTTGAAAAATAT 2384  
Db 2406 GCCTGTCAGCCTTGCTGATCGACTTATTGGGGTGACAACAGATATGACTCTGGATGGAAT 2465  
QY 2385 TAATATAAAGAAAGATATGTTCCCACTTCCCAAGCCAAATACCAAGATGTGCATTTCTT 2444  
Db 2466 CACCTCCCCAGCTGAACCTTTTCCACCTGGAGTCTTGGGAATACCGACGTGATCTTCTT 2525  
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTTAATGAGCCGATCAACTGCTGTGAAAAAT 2504  
Db 2526 TTATAGGTCCAATGATGTGACCCAGTCCCTGCAGTTCTGGGAGATCAACACCATCCGCGT 2585  
QY 2505 GAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCAAGCAAGTCCCGAGC 2564  
Db 2586 CAGGTGCAGTCCACAGAAAACCTGTCCCTGGAAAGTTTGCTGCTGCCAGGAACGTGCTCAGA 2645  
QY 2565 AGGTACCTGTGATGGGTGTACGTTCTATTCTCTGTGGGAGAGTGCTGAAGCTTGCCCTCT 2624  
Db 2646 TGGGACCTGTGATGGCTGCAACTTCCACTTCCCTGTGGGAGAGCGCGGCTGCTTGCCCGCT 2705  
QY 2625 GTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCCTGCAAGAGAGAGGATTTCAGGAAAC 2684  
Db 2706 CTGCTCAGTGGCTGACTACCATGCTATCGTCAAGCAGCTGTGTGGCTGGGATCCAGANGAC 2765  
QY 2685 CTTGTATGTGGAATGAACCTAAATGGTGCATTAAGGAATTTCTTTGCCTGAGAAAAA 2744  
Db 2766 TACTTACGTGTGNCAGAAACCCCAAGCTATGCTCTGGTGGCATTTCTCTGCCTGAGCAGAG 2825  
QY 2745 GTTGGCAACTGTGAAACGGTTGACTTTTGGCTGAAGGTGGGAGCGGCTGTGGGAGCTTT 2804  
Db 2826 AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG 2885  
QY 2805 TACTGCCGTTTGTGGTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATCAAAAGAAAAA 2864  
Db 2886 TACTGCCATCTGCTCACCGTCTTGACCTGCTACTTTTGGAAAAAGAAATCAAAACTAGA 2945  
QY 2865 GAAGACCATTTTGAATCTGTTCA 2887  
Db 2946 GTACAAAGTACTCCAAGCTGGTGA 2968

US-10-176-921-37  
; Sequence 37, Application US/10176921  
; Publication No. US20030027276A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C288  
; CURRENT APPLICATION NUMBER: US/10/176,921  
; CURRENT FILING DATE: 2002-06-20  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 37  
; LENGTH: 3501  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 2762, 2778  
; OTHER INFORMATION: unknown base  
US-10-176-921-37

Query Match 26.6%; Score 901; DB 9; Length 3501;  
Best Local Similarity 60.0%; Pred. No. 2.5e-241;  
Matches 1634; Conservative 0; Mismatches 1062; Indels 27; Gaps 7;

QY 171 GCTTCCTCCTTGCCAGGAGAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGG 230  
Db 267 GCTTCATGCCTGCAAAAGAGTCTGAGTACCACTATGAGTACACGGCGTGTGACAGCACGGG 326  
QY 231 CTCAGGTGGAGAGTTGCCATTCCAAATTTGCAGTGGACTGCTCTGSCCTGCCTGACCC 290  
Db 327 TTCCAGGTGGAGGGTCCCGTCCCGCATACCCCGGCCCTGTGCACCAGCCTGTCTGACCC 386  
QY 291 AGTGAGAGGCAAAAGATGCACCTTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAA 350  
Db 387 CGTCAAGGGCACCGAGTGCTCCTTCTCTGCAACGCCGGGAGTTTCTGGATATGAAGGA 446  
QY 351 CCAGGTATGCAGTAAGTGTGGTGAAGGCACCTATTCTTGGCAGTGGCATCAAATTTGA 410  
Db 447 CCAGTCATGTAAGCCATGCGCTGAGGCGCGCTACTCCCTCGGCACAGGCATTCGGTTTGA 506  
QY 411 TGAATGGGATGAATTGCCGGCAGGATTTCTAAACATCGCAACATTCATGGACACTGTGGT 470  
Db 507 TGAGTGGGATGAGTGCCTCCATGGCTTGCCAGCCTCTCAGCCACACATGGAGCTGGATGA 566  
QY 471 GGGCCCTTCTGACAGCAGGCCAGCGGCTGTAAACAACCTCTTCTTGGATCCCTCGTGGAAA 530  
Db 567 CAGTGCCTGAGTCCA--CCGGGAACTGTACTTCTGTCGAAGTGGGTTCCCGGGGGCGA 623  
QY 531 CTACATAGAATCTAATCGTGTGATGACTGCACGGTGTCTTTTGATCTATGCTGTGCACCTTAA 590  
Db 624 CTACATCGCCTCCAACACGGACGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 683  
QY 591 GAAGTCAGGCTATGCTTCTTTTGTAGTACCAGTATGTCGACAAACAACATCTTCTTTGAGTT 650  
Db 684 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 743

QY 651 CTTTATTCAAATGATCAGTGCAGGAGATGGACACCACCACCTGACAAGTGGGTAAACT 710  
Db 744 TTTTCGTGAGAATGACCAGTGCAGGCCAATGCAGATGACTC--CAGGTGGATGAAGAC 800  
QY 711 TACAGACAATGGAGAATGGGGCTCTCATTTCTGTAATGCTGAATCAGGCACAAACATACT 770  
Db 801 CACAGAGAA--AGGATGGGAATTCACACAGTGTGGAGCTAAATCGAGGCAATAATGTCCT 857  
QY 771 CTACTGGAGAACTACAGGCATCCTTATGGGTTCTTAAGCGGTCAAGCCTGTGCTGTTAAA 830  
Db 858 CTATTGGAGAAACCAAGCCTTCTCAGTATGGACCAAAAGTACCCAAAGCCTGTGCTGGTGAG 917  
QY 831 AAATATCAACAATGAAGGGGTGGGTGACATACAGAAATGTTTTCCTTGAAGCCAGGCAC 890  
Db 918 AAACATTGCCATAACAGGGGTGGCTGACACTTCAGAAATGCTTCCCTTGCAAACCTGGCAC 977  
QY 891 ATTACGCAACAAACCAAGGTTCAATCAACTGCCAGGTGTGTCACAGAAACACCTATTCTGA 950  
Db 978 GTATGCAGACAAGCAGGGCTCCTCTTCTGCAAACTTGTGCCAGCCAACTCTTATTCAAA 1037  
QY 951 GAAAGGAGCCAAAGAATGTATAAGGTGTAAA--GACGACTCTCAATTTTCAGGATCCAG 1007  
Db 1038 TAAAGGAGAAACTTCTTGCCACCAGTGTGACCCTGACAAATACTCAGAGAAAGGATCTTC 1097  
QY 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAGACTATTTCCAGATCCATACTCCATG 1067  
Db 1098 TTCTGTACGTGCGCCCAAGCTTGACAGACAAAGATTATTTCTACACACACACGGCCTG 1157  
QY 1068 TGATGAAGAAAGGAAGACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGGA 1127  
Db 1158 CGATGCCAACGGAGAGACACAACATCATGTACAAATGGGCCAAGCCGAAATCTGTAGCGA 1217  
QY 1128 GGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAAAGGATTGTCCGCCCTTG 1187  
Db 1218 GGACCTTGAGGGGGCAGTGAAGCTGCCTCTGCTGTGAGAGACCCACTGCCCAACCTG 1277  
QY 1188 CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCCTCCTGGAACATT 1247  
Db 1278 CAACCCAGGCTCTCTCAAACCCAACACAGACACCTGCCAGCCCTGCCCATATGTTCCCTA 1337  
QY 1248 TTCAGATGGAACCAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTT 1307  
Db 1338 CTCCAATGG--CTCAGACTGTACCCTGTCCTGTCAGGGACTGAACCTGCTGTGGGATT 1394  
QY 1308 TGAATATAAATGGTGGAAATGTCCTTCTTGGCAACATGAAAACTTCTCTGCTTCAATGTTGG 1367  
Db 1395 TGAATACAATGGTGGAAACACGCTGCCACACAACATGGAAACGACCGTCTCAGTGGGAT 1454  
QY 1368 GAATTCAAAGTCGGATGGAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427  
Db 1455 CAACTTCGAGTACAAAGGGCATGACAGGCTGGGAGGTGGCTGGTGATCACATTTACACAGC 1514  
QY 1428 GGCTGGAGGTTCTGACAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAACC 1487  
Db 1515 TGCTGGAGCCTCAGACAATGACTTCATGATTTCTCACTCTGCTGTTGTGGCAGGATTTAGACC 1574  
QY 1488 ACCAACATCTATGACTGGAGCCAC--GGGTTCTGAACCTAGGAAGATAACATTTGTCTT 1544  
Db 1575 TCCGCAGTCGGTGATGGCAGACACAGAGAATAAAGAGGTGGCCAGAATCACATTTGTCTT 1634  
QY 1545 TGAGACCCCTCTGTCAGCTGACTGTGTTTTGTACTTTCATGGTGGATATTAATAGAAAAAG 1604  
Db 1635 TGAGACCCCTCTGTCTGTGAACGTGTGAGCTTACTTCTATGGTGGGTGGAATTTAGGAC 1694  
QY 1605 TACAAATGTGGTGAATCGTGGGGTGGAAACCAAGAAACAAAGCTTACACCCCATATCAT 1664  
Db 1695 CAACACTCCTGTGGAGACGTGGAAAGGTTCCAAAGGCAACACAGTCTCTATACCTACATCAT 1754  
QY 1665 CTTCAAGAATGCAACTTTTACATTTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGA 1724  
Db 1755 TGAGGAGAACACTACCACGAGCTTCACCTGGGCTTCCAGAGGACCACCTTTTCATGAGGC 1814

QY 1725 TAATAGACGGTTTCATCAATGACATGGTGAAGATTTTATTCTATCAGAGCCACTAATCCAGT 1784  
Db 1815 AAGCAGGAAGTACACCAATGACGTGGCAAGATCTACTCCATCAATGTCAACCAATGTTAT 1874  
QY 1785 TGATGGGGTGGCGTCCCTCATGCCGTGCCCTGTGCCCTCGGTTCTGAACAGTCGGGTTCATC 1844  
Db 1875 GAATGGCGTGGCCTCCTACTGCCGTCCCCTGTGCCCTAGAACCTCTGATGTGGGCTCCTC 1934  
QY 1845 GTGTGTCCCCTGCCCTCCAGGCCACTACATTTGAGAAAGAAACCAACCAAGTGCAGAGAAATG 1904  
Db 1935 CTGCACCTCTTGTCTGTGTTACTATATATTGACCGAGATTAGAACCTGCCACTCCTG 1994  
QY 1905 TCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTCATG 1964  
Db 1995 CCCCCCTAACACAATCTGAAAGCCCAACAGCTTATGTGTCCAGGCTGTGTGCCCTG 2054  
QY 1965 CGGGCTGGGAGTAAACAAACANTCAGGACCATTTCGGTTTGTATAGTACTGCTTTCTA 2024  
Db 2055 TGGTCCAGGGACCAAGAACACAAAGATCCACTCTCTGTGTACAAATGATTGCACCTTCTC 2114  
QY 2025 CCATGAAAAAGAAATCAGATTTTGCACATATGACTTTTAGCAACCTCAGCAGTGTGGGCTC 2084  
Db 2115 ACGCAACACTCCAAACCAAGGACTTCAACTACAACTTCTCCGCTTGGCAACACCCGTCAC 2174  
QY 2085 ATTAAATGAATGGCCCCAGCTTACCTCCAAAGGAACAAAAATCTTCCATTTCTTCAATAT 2144  
Db 2175 TCTTGTGTGGGGGCCAAGCTTCACTTCCAAAGGGTTGAAATACTTCCATCATTACCT 2234  
QY 2145 CAGTTTATGTGGGCATGAGGGGAAGAAAGATGGCTCTCTGTACCAACAAATATACAGACT 2204  
Db 2235 CAGTCTCTGTGGAACCCAGGGTAGGAAAAATGTCTGTGTGACCGGCAAAATGTCACTGACCT 2294  
QY 2205 TACAGTAAAAAGAAATAGTGGCAGGTCAGATGATTACACAAATTTGGTAGGGCATTTGT 2264  
Db 2295 CCGGATTCCTGAGGGTGAGTCAAGG-----TTCTCCAATCTATCACAGCCTACCT 2345  
QY 2265 ATGCCAGTCAACAATTTATTCCTTCTGAAAGTAAGGGTTTCCGAGCAGCCTTATCATCACA 2324  
Db 2346 CTGCCAGGCAGTCACTATCCCCCAGAGGTTGACAGGCTACAAAGCCGGGTTTCTCTCACA 2405  
QY 2325 ATCCATCATTTCTGGCAGATACATTCATAGGAGTCACAGTTGAAACCAACATTTGAAAAATAT 2384  
Db 2406 GCCTGTTCAGCCTTGTGTGATCGACTTATTGGGGTGACAACAGATATGACTCTGGATGGAAT 2465  
QY 2385 TAATATAAAGAAAGATATGTTCCAGTTCCCAACCAAGCCAAATACCAGATGTGCATTTCTT 2444  
Db 2466 CACCTCCCAAGCTGAACCTTTTCCACTGGAGTCTCTTGGGAATACCGGACGTGATCTTCTT 2525  
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTGTATTAATAGGCCGATCAACTGCTGTGAAAAAT 2504  
Db 2526 TTATAGSTCCAATGATGTGACCCAGTCTCTGAGTTCTGGGAGATCAACCAACCATCCGCT 2585  
QY 2505 GAGGTGTAATCTTAAATCTGGAGCAGGAGTGATTTTCACTCCCAAGCAAGTGCACAGC 2564  
Db 2586 CAGGTGCAGTCCACAGAAAACTGTCCCTGGAAGTTTGTGCTGCTGCCAGGAACGTGCTCAGA 2645  
QY 2565 AGGTACCTGTGATGGGTGTACGTTCTATTTCTGTGGGAGAGTGTGAAAGCTTGCCCTCT 2624  
Db 2646 TGGGACCTGTGATGGCTGCAACTTCCACTTCTCTGTGGGAGAGCGGCTGTGCCCGCT 2705  
QY 2625 GTGTACGGAGCATGACTTCCATGAGATTCAGGGGAGCCTGCAAGAGAGGATTTTCAGGAAAC 2684  
Db 2706 CTGCTCAGTGGCTGACTACCATGCTATCTGCTCAGCAGCTGTGTGGCTGGGATCCAGANGAC 2765  
QY 2685 CTTGTATGTGTGGAATGAACCTAAATGGTGCATTTAAAGGAATTTCTTTCCTGAGAAAAA 2744  
Db 2766 TACTTACGTGTGNCGAGAACCCCAAGCTATGCTCTGTGGCATTTCTGTGCTGTGAGCAGAG 2825  
QY 2745 GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGGTGGGAGCCGCTGTGGGAGCTTT 2804  
Db 2826 AGTCACCATCTGCAAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG 2885  
QY 2805 TACTGCCGTTTGTGCTGGCTCTGACCTGCTGACTTCTGGAAAAAGAAATCAAAAAAGAAAA 2864



Db	2886	TACTGCCATCCTGCTCACCGTCTTGACCTGCTACTTTTGGAAAAAGAAATCAAAAAC	2945
QY	2865	GAAGACCATTTTGAATCTGTTCA	2887
Db	2946	GTACAAGTACTCCCAAGCTGGTGA	2968

Search completed: May 12, 2003, 13:07:24  
Job time : 541.238 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.  
  
OM nucleic - nucleic search, using sw model  
  
Run on: May 12, 2003, 00:15:42 ; Search time 173.258 Seconds  
(without alignments)  
6000.498 Million cell updates/sec  
  
Title: US-10-073-333A-1  
Perfect score: 3390  
Sequence: 1 atgctgttcgcgcgccgggg.....ggtataaaaaaaaaaaaaa 3390  
  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
  
Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	493.2	14.5	1129	4	US-09-227-357-40
2	43.2	1.3	1120	3	US-09-030-613-1
3	43.2	1.3	1120	4	US-09-451-905-1
4	42	1.2	1209	4	US-09-105-537-21
5	42	1.2	13613	4	US-09-105-537-3
6	42	1.2	38506	3	US-09-320-878-19
7	41.8	1.2	1743	3	US-08-665-259-20
8	41.8	1.2	1743	3	US-08-762-500-20
9	41.8	1.2	1974	3	US-08-762-500-78
10	41.8	1.2	6803	3	US-08-665-259-19
11	41.8	1.2	6803	3	US-08-762-500-19
12	41.2	1.2	2247	2	US-08-524-828-1
13	41.2	1.2	2247	2	US-08-975-114A-1
14	40.8	1.2	16389	4	US-09-741-154-3
15	40.2	1.2	4403765	4	US-09-103-840A-2
16	40.2	1.2	4411529	4	US-09-103-840A-1
17	39.8	1.2	1652	3	US-08-758-662-3
18	39.6	1.2	2834	4	US-09-305-384-6
19	39.6	1.2	6235	4	US-09-305-384-5
20	39.6	1.2	6679	4	US-09-305-384-1
21	39.4	1.2	3507	2	US-08-775-009-36
22	38.8	1.1	9521	4	US-08-972-218-2
23	38.4	1.1	1283	1	US-07-885-970A-17
24	38.4	1.1	1283	1	US-08-298-687A-17
25	38.4	1.1	1283	1	US-08-530-797-18
26	38.4	1.1	1283	1	US-08-298-829-17
27	38.4	1.1	1283	2	US-08-787-335-18

Sequence 1, Appli  
Sequence 14, Appl  
Sequence 1, Appli  
Sequence 3, Appli  
Sequence 1, Appli  
Sequence 3, Appli  
Sequence 11, Appl  
Sequence 6, Appli  
Sequence 12, Appl  
Sequence 5, Appli  
Sequence 1, Appli  
Sequence 107, App  
Sequence 107, App  
Sequence 107, App  
Sequence 107, App  
Sequence 107, App  
Sequence 36, Appl  
Sequence 38, Appl

ALIGNMENTS

RESULT 1  
US-09-227-357-40  
; Sequence 40, Application US/09227357  
; Patent No. 6342581  
; GENERAL INFORMATION:  
; APPLICANT: Fischer et al.  
; TITLE OF INVENTION: 123 Human Secreted Proteins  
; FILE REFERENCE: pz010p1  
; CURRENT APPLICATION NUMBER: US/09/227,357  
; CURRENT FILING DATE: 1999-01-08  
; EARLIER APPLICATION NUMBER: PCT/US98/13684  
; EARLIER FILING DATE: 1998-07-07  
; EARLIER APPLICATION NUMBER: 60/051,926  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,793  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,925  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,929  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,803  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,732  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,931  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,932  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,916  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,930  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,918  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,920  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,733  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,795  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,919  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,928  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/055,722  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,723  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,948  
; EARLIER FILING DATE: 1997-08-18



EARLIER APPLICATION NUMBER: 60/055,949  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,953  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,950  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,947  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,964  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/056,360  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,684  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,984  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,954  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/058,785  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,664  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,660  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,661  
EARLIER FILING DATE: 1997-09-12  
NUMBER OF SEQ ID NOS: 672  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 40  
LENGTH: 1129  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1053)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-227-357-40

Query Match 14.5%; Score 493.2; DB 4; Length 1129;  
Best Local Similarity 98.3%; Pred. No. 2.9e-130;  
Matches 509; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 2860 AAAAAGAGACCATTTTGAATCTGTCAACTGAAACCTCAAGATCCCCCAATATATGAA 2919  
Db 446 AAAAAGAGACCATTTTGAATCTGTCAACTGAAACCTCAAGATCCCCCAATATATGAA 505  
QY 2920 GAGACAGTGTGTAG-CCTTGAGACTAATGAACAAAGAAACCTGCTCTAGTTTACAGGA 2978  
Db 506 GAGACAGTGTGTAGCCCTTGAGACTAATGAACAAAGAAACCTGCTCTAGTTTACAGGA 565  
QY 2979 CCATATTTTAGGGTCTCTCATACCTCTCAGATTTGGTGATCTCAGAGGAGGCCAT 3038  
Db 566 CCATATTTTAGGGTCTCTCATACCTCTCAGATTTGGTGATCTCAGAGGAGGCCAT 625  
QY 3039 GCCGCTGAAAGGGGAGGAGATTGAAACATTTGATTGCCCTTATCAGATGGTCAAGTACCT 3098  
Db 626 GCCGCTGAAAGGGGAGGAGATTGAAACATTTGATTGCCCTTATCAGATGGTCAAGTACCT 685  
QY 3099 TGCCAAATAAAGGAACCAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGA 3158  
Db 686 TGCCAAATAAAGGAACCAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGA 745  
QY 3159 AGAGATTTATCTGTATATACACATAACTGAAACCAAGTTTAAGCCCACTCACTGC 3218  
Db 746 AGAGATTTATCTGTATATACACATAACTGAAACCAAGTTTAAGCCCACTCACTGC 805  
QY 3219 TGATGCATGCCATATAATTAATGGGTAACCTTTTATCTTTATGATGCTACATAACAAGT 3278  
Db 806 TGATGCATGCCATATAATTAATGGGTAACCTTTTATCTTTATGACGCTACATAACAAGT 865  
QY 3279 GTGATTTGGAAGGCACATGTGAGCATATGCAATTATGATCCAAATTTATGTTTTCTTTGT 3338  
Db 866 GTGATTTGGAAGGCACATGTGAGCATATGCAATTATGATCCAAATTTATGTTTTCTTTGT 925

QY 3339 TTATATTTTGGGGAAAAATTAATAATTTTAAAGGTAAA 3376  
Db 926 TTATATTTTGGGGAAAAATTAATAATTTTAAAGGTATA 963  
RESULT 2  
US-09-030-613-1  
Sequence 1, Application US/09030613  
Patent No. 6083706  
GENERAL INFORMATION:  
APPLICANT: Florkiewicz, Robert Z.  
APPLICANT: Baird, J. Andrew  
TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/030,613  
FILING DATE: 25-FEB-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6083706tenburg Ph.D., Carol  
REGISTRATION NUMBER: 39,317  
REFERENCE/DOCKET NUMBER: 760100.418C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1120 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-030-613-1

Query Match 1.3%; Score 43.2; DB 3; Length 1120;  
Best Local Similarity 63.5%; Pred. No. 0.039;  
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 4 CTGTTCCGGCCCGGGCGGTACGGGCGAGGGCGTGGGGCGGGCGGGAGGCTCCC 63  
Db 258 CGGGGCGGTGCCCCGAGCGGGTCCGAGCGCGGGCGGGCGGGCGGGAGGCTCCC 317  
QY 64 CGCCGCGGGCGCTCGCCGCCCTGGAGCCCGCCCTGGATTGCTG 107  
Db 318 CGCGGGCGCTCCAGCGGCTCGGGGATCCCGGGCGGGCGGGCGGGCAG 361

RESULT 3  
US-09-451-905-1  
Sequence 1, Application US/09451905  
Patent No. 6306613  
GENERAL INFORMATION:  
APPLICANT: Robert Z. Florkiewicz  
APPLICANT: Andrew Baird  
APPLICANT: Dale E. Warnock  
TITLE OF INVENTION: MODULATORS OF LEADERLESS PROTEIN EXPORT  
TITLE OF INVENTION: AND METHODS FOR IDENTIFYING AND USING THE SAME  
FILE REFERENCE: 200124.402C4  
CURRENT APPLICATION NUMBER: US/09/451,905  
CURRENT FILING DATE: 1999-12-01  
NUMBER OF SEQ ID NOS: 48

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1120
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-451-905-1

Query Match          1.3%; Score 43.2; DB 4; Length 1120;
Best Local Similarity 63.5%; Pred. No. 0.039;
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 4 CTGTTCCGCGCCCGGGGGCGGTACGGGCGAGGGGCTGGGGCGGCCCGCGGAGGCTCCC 63.
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 258 CGGGCGCGTGTCCCGGAGCGGTCTGGAGGCCGGGCGCGGCCCGGGACGCGCGGCTCCC 317

QY 64 CGCCGCGGCGCTCGCCGCCCTGGAGCCCCCGCTGGATTGCTG 107
   | | | | | | | | | | | | | | | | | | | | | |
Db 318 CGGCGCGGCTCCAGCGGCTCGGGGATCCCGGCCGGCCCCGCAG 361

RESULT 4
US-09-105-537-21
; Sequence 21, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-21

Query Match          1.2%; Score 42; DB 4; Length 1209;
Best Local Similarity 62.3%; Pred. No. 0.089;
Matches 66; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 28 CGGGCAGGGGCTGGGGCGCGCGGAGGCTCCCCCGCGGGCGCTCGCGCCCTGG 87
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 821 CGGCGCGCGCGGTGGAGAGACCCCTCCGCTACGACCCCGCGGTGCAGCTCGACGCCCGGG 880

QY 88 AGCCCCGCCTGGATTGCTGGGCGCTCGCCGGTGCCAGGCGG 133
   | | | | | | | | | | | | | | | | | | | | | |
Db 881 TGGTCCGCGGGAGACGAGCTGGCGGGCGCGGCTGCCGCGCGG 926

RESULT 5
US-09-105-537-3
; Sequence 3, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
```

```
US-09-105-537-3

Query Match          1.2%; Score 42; DB 4; Length 13613;
Best Local Similarity 62.3%; Pred. No. 0.44;
Matches 66; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 28 CGGGCAGGGGCTGGGGCGCGCGGAGGCTCCCCCGCGGGCGCTCGCCGCCCTGG 87
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1626 CGGCGCGCGCGGTGGAGGAGACCTCCGCTACGACCCCGCGGTGCAGCTCGACGCCCGGG 1685

QY 88 AGCCCCGCCTGGATTGCTGGGCGCTCGCCGGTGCCAGGCGG 133
   | | | | | | | | | | | | | | | | | | | | | |
Db 1686 TGGTCCGCGGGAGACGAGCTGGCGGGCGCGGCTGCCGCGCGG 1731

RESULT 6
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match          1.2%; Score 42; DB 3; Length 38506;
Best Local Similarity 62.3%; Pred. No. 0.88;
Matches 66; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 28 CGGGCAGGGGCTGGGGCGCGCGGAGGCTCCCCCGCGGGCGCTCGCCGCCCTGG 87
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 35623 CGGCGCGCGCGGTGGAGGAGACCCCTCCGCTACGACCCCGCGGTGCAGCTCGACGCCCGGG 35682

QY 88 AGCCCCGCCTGGATTGCTGGGCGCTCGCCGGTGCCAGGCGG 133
   | | | | | | | | | | | | | | | | | | | | | |
Db 35683 TGGTCCGCGGGAGACGAGCTGGCGGGCGCGGCTGCCGCGCGG 35728

RESULT 7
US-08-665-259-20
; Sequence 20, Application US/08665259
; Patent No. 6028173
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
```

```

; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES.
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,259
; FILING DATE: 17-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IG5-9.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1743 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1740
; US-08-665-259-20

Query Match 1.2%; Score 41.8; DB 3; Length 1743;
Best Local Similarity 59.8%; Pred. No. 0.13;
Matches 70; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

; QY 6 GTTCCGCGCGCGGGGGCGGTACGGGGCAGGGGCTGGGGCGCGCGGAGGCTCCCGG 65
; || ||||| ||| | ||||| ||| | ||||| ||| | ||||| ||| | |||||
; Db 1505 GTAGCGCGCTGTGGTGCCCGCGGGGATGCGGCCCTGGCGCTGCCGCGCTCCCGG 1564

; QY 66 CCGCGGCGCGCTCGCGCGCCCTGGAGCCCCCGCCCTGGATTGCTGGCGGCTCGCCGG 122
; | || | || | || | || | || | || | || | || | || | || | ||
; Db 1565 GCCGCGCTACCTCCTGCTGGGGGCGGGCCTGGAGCCGCGGCTGGGGCGCGGGGG 1621

RESULT 8
US-08-762-500-20
; Sequence 20, Application US/08762500
; Patent No. 6030806
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klingner, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/762,500
; FILING DATE: 09-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/665,259
; FILING DATE: 17-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10469
; FILING DATE: 17-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IG5-9.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1743 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1740
; US-08-762-500-20

Query Match 1.2%; Score 41.8; DB 3; Length 1743;
Best Local Similarity 59.8%; Pred. No. 0.13;
Matches 70; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

; QY 6 GTTCCGCGCGCGGGGGCGGTACGGGGCAGGGGCTGGGGCGCGCGGAGGCTCCCGG 65
; || ||||| ||| | ||||| ||| | ||||| ||| | ||||| ||| | |||||
; Db 1505 GTAGCGCGCTGTGGTGCCCGCGGGGATGCGGCCCTGGCGCTGCCGCGCTCCCGG 1564

; QY 66 CCGCGGCGCGCTCGCGCGCCCTGGAGCCCCCGCCCTGGATTGCTGGCGGCTCGCCGG 122
; | || | || | || | || | || | || | || | || | || | || | ||
; Db 1565 GCCGCGCTACCTCCTGCTGGGGGCGGGCCTGGAGCCGCGGCTGGGGCGCGGGGG 1621

RESULT 9
US-08-762-500-78
; Sequence 78, Application US/08762500
; Patent No. 6030806
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klingner, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```



```

; APPLICATION NUMBER: US/08/762,500
; FILING DATE: 09-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/665,259
; FILING DATE: 17-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10469
; FILING DATE: 17-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IG5-9.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1974 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-762-500-78

```

Query Match	1.28;	Score 41.8;	DB 3;	Length 1974;
Best Local Similarity	59.88;	Pred. NO. 0.14;		
Matches 70;	Conservative	0;	Mismatches 47;	Indels 0;
			Gaps	0;

Qy	6	GTTCGCGCCCGGGGCCCGGTACGGGGCAGGGCTGGGGGGCGGCCGAGGCTCCCCG	65
Db	1714	GTAGCGCGCTGTGGTGCCCCCGGGGATCGGCGCTCGGGCTGCCCGGCGCTGCTCCCCG	1773
Qy	66	CCGCGGGCGCTCGCGGCCCTGGAGCCCCCGCTGGATTGCTGGGCGCTCGCCGG	122
Db	1774	GCCGCGCGCTACCTCCTGCTGGGGGGCGGGCCCTGGAGCCCGGGCTGGGGGCGCGGGGG	1830

RESULT 10  
US-08-665-259-19  
; Sequence 19, Application US/08665259  
; Patent No. 6028173  
; GENERAL INFORMATION:  
; APPLICANT: Landes, Gregory M.  
; APPLICANT: Burn, Timothy C.  
; APPLICANT: Connors, Timothy D.  
; APPLICANT: Dackowski, William R..  
; APPLICANT: Van Raay, Terence J.  
; APPLICANT: Klinger, Katherine W.  
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,  
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME  
; NUMBER OF SEQUENCES: 73  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENZYME CORPORATION  
; STREET: One Mountain Road  
; CITY: Framingham  
; STATE: Massachusetts  
; COUNTRY: United States of America  
; ZIP: 01701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/665,259  
; FILING DATE: 17-JUN-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dugan, Deborah A.  
; REGISTRATION NUMBER: 37,315  
; REFERENCE/DOCKET NUMBER: IG5-9.1  
; TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6803 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-665-259-19

```

```
Query Match      1.2%; Score 41.8; DB 3; Length 6803;
Best Local Similarity 59.8%; Pred. No. 0.32;
Matches 70; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
```

QY	6	GTTCGGCGCCCGGGGCCCGTACGGGGCAGGGGCTGGGGGGCGGCCGGAGGCTCCCCG	65
Db	5298	GTAGCGCGCTGTGGGTGCCCCCGGGGATCGGGCCTCGGGCTGCCCGCGCTGCTCCCCG	5357
QY	66	CCGGGGGCGCTCGCGCCCTGGAGCCCCCGCCTGGATTGCTGCTGGGCGCTCGCCCG	122
Db	5358	GCCGCGCGTACCTCTGCTGGGGGGCGGGCCCTGGAGCCGCGGCTGGGGGGCGCGGGG	5414

```

RESULT 11
US-08-762-500-19
; Sequence 19, Application US/08762500
; Patent No. 6030806
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/762,500
; FILING DATE: 09-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/665,259
; FILING DATE: 17-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10469
; FILING DATE: 17-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IG5-9.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6803 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```



```
; SEQ ID NO 3
; LENGTH: 16389
; TYPE: DNA
; ORGANISM: Human
US-09-741-154-3

Query Match      1.2%; Score 40.8; DB 4; Length 16389;
Best Local Similarity 58.1%; Pred. No. 1.1;
Matches 72; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 32 GCAGGGGCTGGGGGCGCGCGGAGGCTCCCGCGGGGGCGCTCGCCGCCCTGGAGCC 91
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6006 GCGGTCCTAGCGCGCGCGCACTGCGGTCTCTCCGCGCGCGCGCGCGCGCC 5947

QY 92 CCGCCTGGATTGCTGCTGGCGCTCGCCGGCTGCCAGCGGCGCTGGGCTGGGACCTGC 151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5946 CTCCCCGCGCGCTCGCGCGCGCGCGCGCGCTCCCGCGCGCGCGCGCGCGGG 5887

QY 152 CCTC 155
    |
Db 5886 GCGC 5883

RESULT 15
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      1.2%; Score 40.2; DB 4; Length 4403765;
Best Local Similarity 52.7%; Pred. No. 42;
Matches 87; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 22 CCGGTACGGGGCAGGGGCTGGGGCGCGCGGAGGCTCCCGCGGGGGCGCTCGCCG 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3949006 CCGTCAGTGGCAGCGGTAGTGGCGCGCGGTGCTACCGCGATTCGCGGCCAC 3949065

QY 82 CCCTGGAGCCCCCGCTGGATTGCTGCTGGGCGCTCGCGGCTGCCAGCGGCGCTGGGCT 141
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3949066 ACCTGGCGCACGATGTCGATGATTTCGCGGTGGAGCCAGCGGCTTGGGAAACCGCTGC 3949125

QY 142 GGGACCTGCCCTCCTCCTCCAGCGCGCGCTTCCTCCTTGCCAG 186
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3949126 CCGTACCAACCTCGACCCACTGCGGACCGGACACGCGGAGCCCG 3949170

Search completed: May 12, 2003, 11:08:26
Job time : 11856.3 secs
```





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 05:40:13 ; Search time 546.762 Seconds  
(without alignments)  
8082.236 Million cell updates/sec

Title: US-10-073-333A-3  
Perfect score: 3556  
Sequence: 1 atgctgttcgcgcgcgggg.....ggtaaaaaaaaaaaaaaa 3556

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3556	100.0	3556	9 US-10-140-164-3	Sequence 3, Appli
2	3214	90.4	3390	9 US-10-140-164-1	Sequence 1, Appli
3	1499.6	42.2	1737	9 US-10-002-050-19	Sequence 19, Appli
4	1499.6	42.2	1737	9 US-10-002-304-19	Sequence 19, Appli
5	1499.6	42.2	1737	12 US-10-003-152-19	Sequence 19, Appli
6	1182.6	33.3	1508	9 US-10-002-050-9	Sequence 9, Appli
7	1182.6	33.3	1508	9 US-10-002-304-9	Sequence 9, Appli
8	1182.6	33.3	1508	12 US-10-003-152-9	Sequence 9, Appli
9	987.8	27.8	3501	9 US-10-028-072-37	Sequence 37, Appli
10	987.8	27.8	3501	9 US-10-121-049-37	Sequence 37, Appli
11	987.8	27.8	3501	9 US-10-123-904-37	Sequence 37, Appli
12	987.8	27.8	3501	9 US-10-140-470-37	Sequence 37, Appli
13	987.8	27.8	3501	9 US-10-175-746-37	Sequence 37, Appli
14	987.8	27.8	3501	9 US-10-176-918-37	Sequence 37, Appli
15	987.8	27.8	3501	9 US-10-176-921-37	Sequence 37, Appli
16	987.8	27.8	3501	9 US-10-137-865-37	Sequence 37, Appli
17	987.8	27.8	3501	9 US-10-140-474-37	Sequence 37, Appli
18	987.8	27.8	3501	9 US-10-142-431-37	Sequence 37, Appli
19	987.8	27.8	3501	9 US-10-143-114-37	Sequence 37, Appli

20	987.8	27.8	3501	9 US-10-140-002-37	Sequence 37, Appli
21	987.8	27.8	3501	9 US-10-142-419-37	Sequence 37, Appli
22	987.8	27.8	3501	9 US-10-123-262-37	Sequence 37, Appli
23	987.8	27.8	3501	9 US-10-142-423-37	Sequence 37, Appli
24	987.8	27.8	3501	9 US-10-121-050-37	Sequence 37, Appli
25	987.8	27.8	3501	9 US-10-141-755-37	Sequence 37, Appli
26	987.8	27.8	3501	9 US-10-143-032-37	Sequence 37, Appli
27	987.8	27.8	3501	9 US-10-123-108-37	Sequence 37, Appli
28	987.8	27.8	3501	9 US-10-123-236-37	Sequence 37, Appli
29	987.8	27.8	3501	9 US-10-123-261-37	Sequence 37, Appli
30	987.8	27.8	3501	9 US-10-140-921-37	Sequence 37, Appli
31	987.8	27.8	3501	9 US-10-140-928-37	Sequence 37, Appli
32	987.8	27.8	3501	9 US-10-121-045-37	Sequence 37, Appli
33	987.8	27.8	3501	9 US-10-123-292-37	Sequence 37, Appli
34	987.8	27.8	3501	9 US-10-123-903-37	Sequence 37, Appli
35	987.8	27.8	3501	9 US-10-124-819-37	Sequence 37, Appli
36	987.8	27.8	3501	9 US-10-124-822-37	Sequence 37, Appli
37	987.8	27.8	3501	9 US-10-140-925-37	Sequence 37, Appli
38	987.8	27.8	3501	9 US-10-160-498-37	Sequence 37, Appli
39	987.8	27.8	3501	9 US-10-121-041-37	Sequence 37, Appli
40	987.8	27.8	3501	9 US-10-121-043-37	Sequence 37, Appli
41	987.8	27.8	3501	9 US-10-121-047-37	Sequence 37, Appli
42	987.8	27.8	3501	9 US-10-123-215-37	Sequence 37, Appli
43	987.8	27.8	3501	9 US-10-123-902-37	Sequence 37, Appli
44	987.8	27.8	3501	9 US-10-123-908-37	Sequence 37, Appli
45	987.8	27.8	3501	9 US-10-123-909-37	Sequence 37, Appli

ALIGNMENTS

RESULT 1  
US-10-140-164-3  
; Sequence 3, Application US/10140164  
; Publication No. US20030072736A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker et al.  
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16  
; FILE REFERENCE: PF514C1  
; CURRENT APPLICATION NUMBER: US/10/140,164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 3556  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(3084)  
US-10-140-164-3

Query Match 100.0%; Score 3556; DB 9; Length 3556;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3556; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATGCTGTCCGGCGCGGGGGGGGGTACGGGGCAGGGGCTGGGGCGCGCGGAGGCT 60

Db 1 ATGCTGTTCCGCGCGCGGGGCGCGGTACGGGGCAGGGGCTGGGGCGCGCGGAGGCT 60  
QY 61 CCCCGCGCGGCGCTGCGCCGCGCCTGGAGCCCGCGCTGGATTTGCTGCTGGCGCTCGCC 120  
Db 61 CCCCGCGCGGCGCTGCGCGCGCCTGGAGCCCGCGCTGGATTTGCTGCTGGCGCTCGCC 120  
QY 121 GGCTGCCAGGCGCGCTGGGACCTGCCCTCCTCCTCCAGCCGCGCGCTTCCT 180  
Db 121 GGCTGCCAGGCGCGCTGGGACCTGCCCTCCTCCTCCAGCCGCGCGCTTCCT 180  
QY 181 TGCCAGGAGAAAGATTATCACTTTGAATATACGGAAATGTGATAGCAGTGGCTCCAGGTG 240  
Db 181 TGCCAGGAGAAAGATTATCACTTTGAATATACGGAAATGTGATAGCAGTGGCTCCAGGTG 240  
QY 241 AGAGTTGCCATTCCAAATTCGCAGTGGACTGCTCGCCCTGCCCTGACCCAGTGATGC 300  
Db 241 AGAGTTGCCATTCCAAATTCGCAGTGGACTGCTCGCCCTGCCCTGACCCAGTGATGC 300  
QY 301 AAAGAAATGCACCTTTCTCCTGCTGCTTCTGGAGAGTATCTAGAAATGAAGAACCGATGC 360  
Db 301 AAAGAAATGCACCTTTCTCCTGCTGCTTCTGGAGAGTATCTAGAAATGAAGAACCGATGC 360  
QY 361 AGTAAGTGTGGTGAAGGCACCTATTCTTGGGCAGTGGCATCAAAATTTGATGAATGGGAT 420  
Db 361 AGTAAGTGTGGTGAAGGCACCTATTCTTGGGCAGTGGCATCAAAATTTGATGAATGGGAT 420  
QY 421 GAATGCCGCGCAGGATTTTCTAAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCT 480  
Db 421 GAATGCCGCGCAGGATTTTCTAAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCT 480  
QY 481 GACAGCAGGCCAGACGGCTGTAAACACTCTTCTTGGATCCCTCGTGGAAACTACATAGAA 540  
Db 481 GACAGCAGGCCAGACGGCTGTAAACACTCTTCTTGGATCCCTCGTGGAAACTACATAGAA 540  
QY 541 TCTAATCGTGATGACTGCACGGTGTCTTGTATCTATGCTGTCACCTTAAGAAAGTCAGGC 600  
Db 541 TCTAATCGTGATGACTGCACGGTGTCTTGTATCTATGCTGTCACCTTAAGAAAGTCAGGC 600  
QY 601 TATGCTCTCTTTGAGTACCAGTATGTCGACAAACACATCTCTTTGAGTCTTTATTCAA 660  
Db 601 TATGCTCTCTTTGAGTACCAGTATGTCGACAAACACATCTCTTTGAGTCTTTATTCAA 660  
QY 661 AATGATCAGTCCAGGAGATGGACACCACCTGACAAAGTGGTAAACCTTACAGACAAT 720  
Db 661 AATGATCAGTCCAGGAGATGGACACCACCTGACAAAGTGGTAAACCTTACAGACAAT 720  
QY 721 GGAGAAATGGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATCTCTACTGGAGA 780  
Db 721 GGAGAAATGGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATCTCTACTGGAGA 780  
QY 781 ACTACAGGCATCCTTATGGGTTTAAGCGGTCAAGCCTGTGCTGTTAAATAATATCACA 840  
Db 781 ACTACAGGCATCCTTATGGGTTTAAGCGGTCAAGCCTGTGCTGTTAAATAATATCACA 840  
QY 841 ATTGAAGGGTGGGTACACATAGAAATGTTTCTTCCAGCCAGGCACATTCAGCAAC 900  
Db 841 ATTGAAGGGTGGGTACACATAGAAATGTTTCTTCCAGCCAGGCACATTCAGCAAC 900  
QY 901 AAACCAAGTTCATTCAACTGCCAGTGTGTCCAGAAACACCTATTCTGAGAAAGGAGCC 960  
Db 901 AAACCAAGTTCATTCAACTGCCAGTGTGTCCAGAAACACCTATTCTGAGAAAGGAGCC 960  
QY 961 AAAGAAATGTAAAGGTGTAAGACGAGCTCTCAATTTTTCAGGATCCAGTGATACAGAG 1020  
Db 961 AAAGAAATGTAAAGGTGTAAGACGAGCTCTCAATTTTTCAGGATCCAGTGATACAGAG 1020  
QY 1021 CGCCCTCCCTGTACCACAAAGACTATTCCAGATCCATCTCCATGTGATGAAGAAGGA 1080  
Db 1021 CGCCCTCCCTGTACCACAAAGACTATTCCAGATCCATCTCCATGTGATGAAGAAGGA 1080  
QY 1081 AAGACACAGATAATGTACAAGTGGATAGAGCCCCAAAATCTGCCGGGAGGATCTCACAGAT 1140  
Db 1081 AAGACACAGATAATGTACAAGTGGATAGAGCCCCAAAATCTGCCGGGAGGATCTCACAGAT 1140

Db 1081 AAGACACAGATAATGTACAAGTGGATAGAGCCCCAAAATCTGCCGGGAGGATCTCACAGAT 1140  
QY 1141 GCTATTAGATTGCCCTTCTTGAGAGAAAGGATTTGTCGCGCTTGCACACCTCGATTT 1200  
Db 1141 GCTATTAGATTGCCCTTCTTGAGAGAAAGGATTTGTCGCGCTTGCACACCTCGATTT 1200  
QY 1201 TATAACAATGGATCATCTTCTTGCCATCCCTGTCTCCTCGGAAACATTTTCAGATGGAACC 1260  
Db 1201 TATAACAATGGATCATCTTCTTGCCATCCCTGTCTCCTCGGAAACATTTTCAGATGGAACC 1260  
QY 1261 AAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAATGG 1320  
Db 1261 AAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAATGG 1320  
QY 1321 TGGAAATGCTCCTTCCCTGGCAACATGAAAACTTCCCTGCTTCAATGTTGGGAATTCAAAAGTGC 1380  
Db 1321 TGGAAATGCTCCTTCCCTGGCAACATGAAAACTTCCCTGCTTCAATGTTGGGAATTCAAAAGTGC 1380  
QY 1381 GATGGAATGAATGGTTGGGAGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGTTCCT 1440  
Db 1381 GATGGAATGAATGGTTGGGAGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGTTCCT 1440  
QY 1441 GACAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATG 1500  
Db 1441 GACAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATG 1500  
QY 1501 ACTGGAGCCACGGGTTCTGAACTAGGAAGAAATAACATTTGTCTTTGAGACCCCTCTGTTC 1560  
Db 1501 ACTGGAGCCACGGGTTCTGAACTAGGAAGAAATAACATTTGTCTTTGAGACCCCTCTGTTC 1560  
QY 1561 GCTGACTGTGTTTGTACTTCTGCTGGGATATTAAAGAAAAGTACAAATGTGGTAGAA 1620  
Db 1561 GCTGACTGTGTTTGTACTTCTGCTGGGATATTAAAGAAAAGTACAAATGTGGTAGAA 1620  
QY 1621 TCGTGGGTGGAACCAAGAAAACAAAGCTTACACCCATATCATCTTCAAGANTGCAACT 1680  
Db 1621 TCGTGGGTGGAACCAAGAAAACAAAGCTTACACCCATATCATCTTCAAGANTGCAACT 1680  
QY 1681 TTTTACATTTACATGGGCAATCCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTCAJC 1740  
Db 1681 TTTTACATTTACATGGGCAATCCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTCAJC 1740  
QY 1741 AATGACATGGTGAAGATTTATTTCTATCACAGCCACTAATGCAGTTGATGGGTGGCTCC 1800  
Db 1741 AATGACATGGTGAAGATTTATTTCTATCACAGCCACTAATGCAGTTGATGGGTGGCTCC 1800  
QY 1801 TCATGCCGTGCCCTGTGCCCTCGGTTCTGAAACAGTCCGGTTTCATCGTGTCCCTGCCCT 1860  
Db 1801 TCATGCCGTGCCCTGTGCCCTCGGTTCTGAAACAGTCCGGTTTCATCGTGTCCCTGCCCT 1860  
QY 1861 CCAGGCCACTACATTTGAGAAAGAAACCAACCAAGGATGTCCACCTGACACCTAC 1920  
Db 1861 CCAGGCCACTACATTTGAGAAAGAAACCAACCAAGGATGTCCACCTGACACCTAC 1920  
QY 1921 CTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTCATGCGGGCTGGGAGTAAA 1980  
Db 1921 CTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTCATGCGGGCTGGGAGTAAA 1980  
QY 1981 AACAAATCAGGACCATTCGGTTTGTCTATAGTGACTGCTTTTCTACCATGAAAAAGNAAAT 2040  
Db 1981 AACAAATCAGGACCATTCGGTTTGTCTATAGTGACTGCTTTTCTACCATGAAAAAGNAAAT 2040  
QY 2041 CAGATTTTGCACATATGACTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGGCCCC 2100  
Db 2041 CAGATTTTGCACATATGACTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGGCCCC 2100  
QY 2101 AGCTTCACCTCCAAAGGAACAAAATACTTCCATTTCTTCAATATACAGTTATGTGGGCAT 2160  
Db 2101 AGCTTCACCTCCAAAGGAACAAAATACTTCCATTTCTTCAATATACAGTTATGTGGGCAT 2160  
QY 2161 GAGGGGAAGAGATGGCTCTCTGTACCAACAATATAACAGACTTTACAGTAAAGAAATA 2220  
Db 2161 GAGGGGAAGAGATGGCTCTCTGTACCAACAATATAACAGACTTTACAGTAAAGAAATA 2220



QY 2221 GTGCGAGGTCAGATGATTACACAAATTTGGTAGGGGCATTGTATGCCAGTCAACAATT 2280  
Db 2221 GTGCGAGGTCAGATGATTACACAAATTTGGTAGGGGCATTGTATGCCAGTCAACAATT 2280  
QY 2281 ATTCTCTCTGAAAGTAAGGGTTCCGAGCAGCCCTTATCATCACAAATCCATCATCTCGCA 2340  
Db 2281 ATTCTCTCTGAAAGTAAGGGTTCCGAGCAGCCCTTATCATCACAAATCCATCATCTCGCA 2340  
QY 2341 GATACATTCATAGGAGTCACAGTTGAAACCACATTGAAAAATATTAAATAAAAAGAGAT 2400  
Db 2341 GATACATTCATAGGAGTCACAGTTGAAACCACATTGAAAAATATTAAATAAAAAGAGAT 2400  
QY 2401 ATGTTCCAGTTCCAAACAGCCAAATACCAGATGTGCATTCTCTTTATAAGTCTTCTACA 2460  
Db 2401 ATGTTCCAGTTCCAAACAGCCAAATACCAGATGTGCATTCTCTTTATAAGTCTTCTACA 2460  
QY 2461 GCAACAACATCTTGTATTAAATGGCCGATCAACTGCTGTGAAAAATGAGGTGTAATCCTACT 2520  
Db 2461 GCAACAACATCTTGTATTAAATGGCCGATCAACTGCTGTGAAAAATGAGGTGTAATCCTACT 2520  
QY 2521 AAATCTGGAGCAGGAGTGATTTCAGTCCCCAGCAAGTGCCCAGCAGGTACCTGTGATGGG 2580  
Db 2521 AAATCTGGAGCAGGAGTGATTTCAGTCCCCAGCAAGTGCCCAGCAGGTACCTGTGATGGG 2580  
QY 2581 TGTACGTTCTATTTCCTGTGGGAGAGTCTGAAGCTTGCCCTCTGTGTACGGAGCATGAC 2640  
Db 2581 TGTACGTTCTATTTCCTGTGGGAGAGTCTGAAGCTTGCCCTCTGTGTACGGAGCATGAC 2640  
QY 2641 TTCCATGAGATTGAGGGAGCCTGCAAGAGAGGATTTTCAGGAACCTTGCAACCTGTGAA 2700  
Db 2641 TTCCATGAGATTGAGGGAGCCTGCAAGAGAGGATTTTCAGGAACCTTGCAACCTGTGAA 2700  
QY 2701 GAACCTAAATGGTGCAATTAAGGAATTTCTTTGCCTGAGAAAAAGTTGGCAACCTGTGAA 2760  
Db 2701 GAACCTAAATGGTGCAATTAAGGAATTTCTTTGCCTGAGAAAAAGTTGGCAACCTGTGAA 2760  
QY 2761 ACGGTTGACTTTTGGCTGAAGGTGGGAGCCGGTGTGGGAGCTTTTACTGCCGTTTGGCTG 2820  
Db 2761 ACGGTTGACTTTTGGCTGAAGGTGGGAGCCGGTGTGGGAGCTTTTACTGCCGTTTGGCTG 2820  
QY 2821 GTGGCTGTGACCTGTCTACTTCTGSRAAAAAGAAATCAAAAACCTGGAATACAAATATTCCAAG 2880  
Db 2821 GTGGCTGTGACCTGTCTACTTCTGSRAAAAAGAAATCAAAAACCTGGAATACAAATATTCCAAG 2880  
QY 2881 TTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCCGGCTGCAGACAGTTGTGCTATC 2940  
Db 2881 TTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCCGGCTGCAGACAGTTGTGCTATC 2940  
QY 2941 ATGGAAGGAGAAGATAATGAAGAGGAGTTGTATATTCCAATAAACAGTCACTACTAGGA 3000  
Db 2941 ATGGAAGGAGAAGATAATGAAGAGGAGTTGTATATTCCAATAAACAGTCACTACTAGGA 3000  
QY 3001 AAACCTCAAATCTTTGGCAACCAGGAAAAAGAGACCATTTTGAATCTGTTCAACTGAAA 3060  
Db 3001 AAACCTCAAATCTTTGGCAACCAGGAAAAAGAGACCATTTTGAATCTGTTCAACTGAAA 3060  
QY 3061 ACCTCAAGATCCCCAAATATATGAAGACAGAGTGTGTAGCCCTTGAGACTAATGAACAAA 3120  
Db 3061 ACCTCAAGATCCCCAAATATATGAAGACAGAGTGTGTAGCCCTTGAGACTAATGAACAAA 3120  
QY 3121 GAAACCTGCTCTAGTTTACAGGACCATATTTTAGGGTCTGTCTCTCATACCTGTACATT 3180  
Db 3121 GAAACCTGCTCTAGTTTACAGGACCATATTTTAGGGTCTGTCTCTCATACCTGTACATT 3180  
QY 3181 GGTGATCTCACAGAGGAGGGCCATGCCGCTGAAAAAGGGAAGGAGATTGAAACATTTGATT 3240  
Db 3181 GGTGATCTCACAGAGGAGGGCCATGCCGCTGAAAAAGGGAAGGAGATTGAAACATTTGATT 3240  
QY 3241 GCCTTATCATGATGGTCAAGTACCTTGCACAAATAAGGAAAGCAAAATGATTTGGGTCTCAA 3300  
Db 3241 GCCTTATCATGATGGTCAAGTACCTTGCACAAATAAGGAAAGCAAAATGATTTGGGTCTCAA 3300

QY 3301 CTGAAGATGAAGCTCAACTCAGGAAGAGATTTATCTGTATATACACATAACTGAAAAACA 3360  
Db 3301 CTGAAGATGAAGCTCAACTCAGGAAGAGATTTATCTGTATATACACATAACTGAAAAACA 3360  
QY 3361 AGTTAAAGCCCAACCAATGCACCTGCTGATGCATGCCATATATAATTAATGGTAACTTTATT 3420  
Db 3361 AGTTAAAGCCCAACCAATGCACCTGCTGATGCATGCCATATATAATTAATGGTAACTTTATT 3420  
QY 3421 CTTTATGATGTCTACATAACAAGTGTGATTTGGAAGGCACATGTGAGCATATGCATTATG 3480  
Db 3421 CTTTATGATGTCTACATAACAAGTGTGATTTGGAAGGCACATGTGAGCATATGCATTATG 3480  
QY 3481 ATCCAAATTATGTTTTCTTTCTTTTATATATTTTGGGGAAAAATTAAATTTTAAAGGTA 3540  
Db 3481 ATCCAAATTATGTTTTCTTTCTTTTATATATTTTGGGGAAAAATTAAATTTTAAAGGTA 3540  
QY 3541 AAAAAAATAAAAAA 3556  
Db 3541 AAAAAAATAAAAAA 3556

RESULT 2  
US-10-140-164-1  
; Sequence 1, Application US/10140164  
; Publication No. US20030072736A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker et al.  
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16  
; FILE REFERENCE: PF514C1  
; CURRENT APPLICATION NUMBER: US/10/140,164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 3390  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2892)  
US-10-140-164-1

Query Match 90.4%; Score 3214; DB 9; Length 3390;  
Best Local Similarity 95.3%; Pred. No. 0;  
Matches 3390; Conservative 0; Mismatches 0; Indels 166; Gaps 1;

QY 1 ATGCTGTTCCGCGCCCCGGGGCGGTACGGGGCAGGGGCTGGGGCGCGCGGAGGCT 60  
Db 1 ATGCTGTTCCGCGCCCCGGGGCGGTACGGGGCAGGGGCTGGGGCGCGCGGAGGCT 60  
QY 61 CCCGCCCGCGGCGCTCGCCGCCCTGGAGCCCCCGCTGGATTGTGCTGGGCGCTCGCC 120  
Db 61 CCCGCCCGCGGCGCTCGCCGCCCTGGAGCCCCCGCTGGATTGTGCTGGGCGCTCGCC 120  
QY 121 GGCTGCCAGCGCGCTGGGCTGGGGACCTGCCCTCCTCCTCCAGCCGCCGCTTCCTCCT 180  
Db 121 GGCTGCCAGCGCGCTGGGCTGGGGACCTGCCCTCCTCCTCCTCCAGCCGCCGCTTCCTCCT 180

QY 181 TGCCAGGAGAAAAGATTATCACCTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGG 240  
Db 181 TGCCAGGAGAAAAGATTATCACCTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGG 240  
QY 241 AGAGTTGCCATTCCAAATCTGCAGTGGACTGCTCTGGCCCTGCCCTGACCCAGTGAGAGGC 300  
Db 241 AGAGTTGCCATTCCAAATCTGCAGTGGACTGCTCTGGCCCTGCCCTGACCCAGTGAGAGGC 300  
QY 301 AAAGAAATGCACATTTCTCCTCTGCTTCTGGAGAGTATCTAGAAATGAAGAACCAGGTATGC 360  
Db 301 AAAGAAATGCACATTTCTCCTCTGCTTCTGGAGAGTATCTAGAAATGAAGAACCAGGTATGC 360  
QY 361 AGTAAGTGTGGTGAAGGCACCTATTCTCTGGGAGTGGCATCAAAATTTGATGAATGGGAT 420  
Db 361 AGTAAGTGTGGTGAAGGCACCTATTCTCTGGGAGTGGCATCAAAATTTGATGAATGGGAT 420  
QY 421 GAATTGCCGGCAGGATTTTCTAACATCGCAACATTCATGGGAGTGGTGGGCCCTTCT 480  
Db 421 GAATTGCCGGCAGGATTTTCTAACATCGCAACATTCATGGGAGTGGTGGGCCCTTCT 480  
QY 481 GACAGCAGGCCAGACGGCTGTAAACAACCTCTTCTTGGATCCCGTGGAAACTACATAGAA 540  
Db 481 GACAGCAGGCCAGACGGCTGTAAACAACCTCTTCTTGGATCCCGTGGAAACTACATAGAA 540  
QY 541 TCTAATCGTGATGACTGCACGGTGCTTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGC 600  
Db 541 TCTAATCGTGATGACTGCACGGTGCTTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGC 600  
QY 601 TATGTCTTCTTTGAGTACCAGTATGTCGACAACAACATCTTCTTTGAGTCTTTATTCAA 660  
Db 601 TATGTCTTCTTTGAGTACCAGTATGTCGACAACAACATCTTCTTTGAGTCTTTATTCAA 660  
QY 661 AATGATCAGTCCCAGGAGATGGACACACCACCTGACAAGTGGGTAAACCTTACAGACAAT 720  
Db 661 AATGATCAGTCCCAGGAGATGGACACACCACCTGACAAGTGGGTAAACCTTACAGACAAT 720  
QY 721 GGAGAAATGGGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATACTCTACTGGAGA 780  
Db 721 GGAGAAATGGGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATACTCTACTGGAGA 780  
QY 781 ACTACAGGCATCCTTATGGGTTCTAAGCGGCTCAAGCCTGTGCTGTTAAATAATATCACA 840  
Db 781 ACTACAGGCATCCTTATGGGTTCTAAGCGGCTCAAGCCTGTGCTGTTAAATAATATCACA 840  
QY 841 ATTGAAGGGTGGCGTACACATCAGAAATGTTTCCCTTGCAAGCCAGGCACATTCAGCAAC 900  
Db 841 ATTGAAGGGTGGCGTACACATCAGAAATGTTTCCCTTGCAAGCCAGGCACATTCAGCAAC 900  
QY 901 AAACCAAGGTTCAATCAACTGCCAGGTGTCTCCAGAAACACCTATTCTGAGAAAGGAGCC 960  
Db 901 AAACCAAGGTTCAATCAACTGCCAGGTGTCTCCAGAAACACCTATTCTGAGAAAGGAGCC 960  
QY 961 AAAGAATGTATAAGGTGTAAAGACGACTCTCAATTTTCAGGATCCAGTGAGTGACAGAG 1020  
Db 961 AAAGAATGTATAAGGTGTAAAGACGACTCTCAATTTTCAGGATCCAGTGAGTGACAGAG 1020  
QY 1021 CGCCCTCCCTGTACCACAAAAGACTATTTCAGATCCATACTCCATGTGATGAAGAAGGA 1080  
Db 1021 CGCCCTCCCTGTACCACAAAAGACTATTTCAGATCCATACTCCATGTGATGAAGAAGGA 1080  
QY 1081 AAGACACAGATAATGTACAAGTGGATAGAGCCCCAAAATCTGCCGGGAGGATCTCACAGAT 1140  
Db 1081 AAGACACAGATAATGTACAAGTGGATAGAGCCCCAAAATCTGCCGGGAGGATCTCACAGAT 1140  
QY 1141 GCTATTAGATTGCCCTTCTTGGAGAGAAGGATTTGTCGCCCTTGCAACCCCTGGATTT 1200  
Db 1141 GCTATTAGATTGCCCTTCTTGGAGAGAAGGATTTGTCGCCCTTGCAACCCCTGGATTT 1200  
QY 1201 TATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTGGAACATTTTCAGATGGAACC 1260  
Db 1201 TATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTGGAACATTTTCAGATGGAACC 1260

QY 1261 AAAGAAATGTAGACCATTGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTTGAATATAAATGG 1320  
Db 1261 AAAGAAATGTAGACCATTGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTTGAATATAAATGG 1320  
QY 1321 TGGAAATGTCTCTTCCCTGGCAACATGAAAACCTTCCCTGCTTCAATGTTGGGAATTCAAAAGTGC 1380  
Db 1321 TGGAAATGTCTCTTCCCTGGCAACATGAAAACCTTCCCTGCTTCAATGTTGGGAATTCAAAAGTGC 1380  
QY 1381 GATGGAATGAATGGTGGGAGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCT 1440  
Db 1381 GATGGAATGAATGGTGGGAGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCT 1440  
QY 1441 GACAAATGATTACCTGATCTTAAACTTGCNTATCCCAGGATTTAAACCCACCAACATCTATG 1500  
Db 1441 GACAAATGATTACCTGATCTTAAACTTGCNTATCCCAGGATTTAAACCCACCAACATCTATG 1500  
QY 1501 ACTGGAGCCACGGGTTCTGAACATAGGAAGAAATAACATTTGTCTTTGAGACCCCTCTGTTCA 1560  
Db 1501 ACTGGAGCCACGGGTTCTGAACATAGGAAGAAATAACATTTGTCTTTGAGACCCCTCTGTTCA 1560  
QY 1561 GCTGACTGTGTTTTGTACTTTTCATGGTGGATATTAAAGAAAAGTACAAAATGTGGTAGAA 1620  
Db 1561 GCTGACTGTGTTTTGTACTTTTCATGGTGGATATTAAAGAAAAGTACAAAATGTGGTAGAA 1620  
QY 1621 TCGTGGGGTGGAAACCAAGAAACAAAGCTTACACCCCATATCATCTTCAAGAAATGCAACT 1680  
Db 1621 TCGTGGGGTGGAAACCAAGAAACAAAGCTTACACCCCATATCATCTTCAAGAAATGCAACT 1680  
QY 1681 TTTTACATTTACATGGGCATTTCCAGAGAACTAAATCAGGGTCAAGATAATAGACGGTTTCATC 1740  
Db 1681 TTTTACATTTACATGGGCATTTCCAGAGAACTAAATCAGGGTCAAGATAATAGACGGTTTCATC 1740  
QY 1741 AATGACATGGTGAAGATTTATTTCTATCACAGCCACTAAATGCAGTTGATGGGTGGCGTCC 1800  
Db 1741 AATGACATGGTGAAGATTTATTTCTATCACAGCCACTAAATGCAGTTGATGGGTGGCGTCC 1800  
QY 1801 TCATGCGGTGCCCTGTGCCCTCGGTTCTGAAACAGTGGGTTTCATCGTGTGCCCTGCCCT 1860  
Db 1801 TCATGCGGTGCCCTGTGCCCTCGGTTCTGAAACAGTGGGTTTCATCGTGTGCCCTGCCCT 1860  
QY 1861 CCAGGCCACTACATTTGAGAAAGAAACCAACCAGTGCAAGGAAATGTCCACCTGACACCTAC 1920  
Db 1861 CCAGGCCACTACATTTGAGAAAGAAACCAACCAGTGCAAGGAAATGTCCACCTGACACCTAC 1920  
QY 1921 CTGTCCATACATCAGGTCTATGGCAACAGGCTTGTATTTCCATCGCGGCCCTGGGAGTAAA 1980  
Db 1921 CTGTCCATACATCAGGTCTATGGCAACAGGCTTGTATTTCCATCGCGGCCCTGGGAGTAAA 1980  
QY 1981 AACAAATCAGGACCATTCGGTTTGTATAGTGACTGCTTTTTCTACCATGAAAAAAGAAAAT 2040  
Db 1981 AACAAATCAGGACCATTCGGTTTGTATAGTGACTGCTTTTTCTACCATGAAAAAAGAAAAT 2040  
QY 2041 CAGATTTTGCACATGACTTTAGCAACCTCAGCAGTGTGGGCTCAATTAATGAATGGCCCC 2100  
Db 2041 CAGATTTTGCACATGACTTTAGCAACCTCAGCAGTGTGGGCTCAATTAATGAATGGCCCC 2100  
QY 2101 AGCTTCACTCCAAAGGAACAAAATACTTCCATTTCTTCAATATCAGTTATGTGGGCAT 2160  
Db 2101 AGCTTCACTCCAAAGGAACAAAATACTTCCATTTCTTCAATATCAGTTATGTGGGCAT 2160  
QY 2161 GAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTTTTACAGTAAAAAGAAATA 2220  
Db 2161 GAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTTTTACAGTAAAAAGAAATA 2220  
QY 2221 GTGGCAGGGTCAGATGATTACACAAATTTGGTAGGGGCAATTTGTATGCCAGTCAACAAAT 2280  
Db 2221 GTGGCAGGGTCAGATGATTACACAAATTTGGTAGGGGCAATTTGTATGCCAGTCAACAAAT 2280  
QY 2281 ATTCTTCTGAAAAGTAAGGGTTTCCGAGCAGCCTTATCATCACATCCATCATTTCTGGCA 2340  
Db 2281 ATTCTTCTGAAAAGTAAGGGTTTCCGAGCAGCCTTATCATCACATCCATCATTTCTGGCA 2340  
QY 2341 GATACATTTCATAGGAGTCACAGTTGAAACCACATTTGAAAAATATTTAATATAAAAAAGAT 2400

||||| 2341 GATACATTCATAGGAGTCACAGTTGAAACCACATTTGAAAAATATTAATATAAAAGAGAT 2400  
QY 2401 ATGTTCCAGTTCCAACAAGCCAAATACCAGATGTGCATTTCTTTTATAAGTCTTCTACA 2460  
Db 2401 ATGTTCCAGTTCCAACAAGCCAAATACCAGATGTGCATTTCTTTTATAAGTCTTCTACA 2460  
QY 2461 GCAACAACATCTTGTTAATGCGCGATCAACTGCTGTGAAAAATGAGGTGTAATCCTACT 2520  
Db 2461 GCAACAACATCTTGTTAATGCGCGATCAACTGCTGTGAAAAATGAGGTGTAATCCTACT 2520  
QY 2521 AAATCTGGACAGGAGTGATTTACGTCCCCAGCAAGTGCCACAGAGTACCTGTGATGGG 2580  
Db 2521 AAATCTGGACAGGAGTGATTTACGTCCCCAGCAAGTGCCACAGAGTACCTGTGATGGG 2580  
QY 2581 TGTACGTTCTATTCTCTGTGGGAGAGTGTGAAGCTTGCCCTCTGTACGGAGCATGAC 2640  
Db 2581 TGTACGTTCTATTCTCTGTGGGAGAGTGTGAAGCTTGCCCTCTGTACGGAGCATGAC 2640  
QY 2641 TTCCATGAGATTGAGGGAGCCTGCAACAGAGGATTTTCAGGAAACCTTGTATGTGTGGAAT 2700  
Db 2641 TTCCATGAGATTGAGGGAGCCTGCAACAGAGGATTTTCAGGAAACCTTGTATGTGTGGAAT 2700  
QY 2701 GAACCTAAATGGTGCATTAAAGGAATTTCTTTCCTGTGAGAAAAAGTTGGCAACCTGTGAA 2760  
Db 2701 GAACCTAAATGGTGCATTAAAGGAATTTCTTTCCTGTGAGAAAAAGTTGGCAACCTGTGAA 2760  
QY 2761 ACGGTTGACTTTGGCTGAAGGTGGGAGCGGTGTGGGAGCTTTTACTGCCGTTTGTCTG 2820  
Db 2761 ACGGTTGACTTTGGCTGAAGGTGGGAGCGGTGTGGGAGCTTTTACTGCCGTTTGTCTG 2820  
QY 2821 GTGGCTCTGACCTGCTACTTCTGGAAGAAATCAAAAACTGGAATACAAAATATTCCAAG 2880  
Db 2821 GTGGCTCTGACCTGCTACTTCTGGAAGAAATCAAAAACTGGAATACAAAATATTCCAAG 2880  
QY 2881 TTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCCGCTGCAGACAGTTGTGCTATC 2940  
Db 2859 ----- 2858  
QY 2941 ATGGAAGGAGAGATAATGAAGAGGAAGTTGTATATTCCAATAAACAGTCACTACTAGGA 3000  
Db 2859 ----- 2858  
QY 3001 AAACCTCAAATCTTTGGCAACCAAGGAAAAAGAGACCATTTTGAATCTGTTCAACTGAAA 3060  
Db 2859 -----GAAAAAGAGACCATTTTGAATCTGTTCAACTGAAA 2894  
QY 3061 ACCTCAAGATCCCCAAATATATGAAGAGACAGTGTGTAGCCCTTGAGACTAATGAACAAA 3120  
Db 2895 ACCTCAAGATCCCCAAATATATGAAGAGACAGTGTGTAGCCCTTGAGACTAATGAACAAA 2954  
QY 3121 GAAACCTGCTCTAGTTTACAGGACCATATTTAGGGTCTGTCTCTCATACCTGTACACATT 3180  
Db 2955 GAAACCTGCTCTAGTTTACAGGACCATATTTAGGGTCTGTCTCTCATACCTGTACACATT 3014  
QY 3181 GGTGATCTCACAGAGGAGGGCCATGCCGCTGAAAAGGGGAGGAGATTGAAACATTTGATT 3240  
Db 3015 GGTGATCTCACAGAGGAGGGCCATGCCGCTGAAAAGGGGAGGAGATTGAAACATTTGATT 3074  
QY 3241 GCGTTATCACATGGTCAAGTACCTTGCCAAATAAAGGAAAGCAAAATGATTGGGTCTCAA 3300  
Db 3075 GCGTTATCACATGGTCAAGTACCTTGCCAAATAAAGGAAAGCAAAATGATTGGGTCTCAA 3134  
QY 3301 CTGAAGATGAAGCTCAACTCAGGAAGAGATTTATCTGTATATACACATAACTGAAAAACCA 3360  
Db 3135 CTGAAGATGAAGCTCAACTCAGGAAGAGATTTATCTGTATATACACATAACTGAAAAACCA 3194  
QY 3361 AGTTTAAGCCCCACCAATGCACTGCTGATGCATGCCATATAATTAATGGGTAACTTTATT 3420  
Db 3195 AGTTTAAGCCCCACCAATGCACTGCTGATGCATGCCATATAATTAATGGGTAACTTTATT 3254  
QY 3421 CTTTATGATGTCTACATAACAAGTGTGATTTGGAGGACACATGTGAGCATATGCATTATG 3480  
|||||



Db 3255 CTTTATGATGTCTACATAACAAGTGTGATTTGGAAGGCACATGTGAGCATATGCATTATG 3314  
QY 3481 ATCCAATTTATGTTTTTCTTTTGTATATTTTGGGAAAAAATTTAAAGGTA 3540  
Db 3315 ATCCAATTTATGTTTTTCTTTTGTATATTTTGGGAAAAAATTTAAAGGTA 3374  
QY 3541 AAAAAAATAAAAAA 3556  
Db 3375 AAAAAAATAAAAAA 3390  
  
RESULT 3  
US-10-002-050-19  
; Sequence 19, Application US/10002050  
; Publication No. US20030032095A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: No. US20030032095A1  
; FILE REFERENCE: 15966-554 Cura-54 CON:S14  
; CURRENT APPLICATION NUMBER: US/10/002,050  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 1737  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (296)..(1687)  
; NAME/KEY: variation  
; LOCATION: (1)..(1737)  
; OTHER INFORMATION: N may be any nucleotide  
US-10-002-050-19

Query Match 42.2%; Score 1499.6; DB 9; Length 1737;  
Best Local Similarity 98.7%; Pred. No. 0;  
Matches 1545; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

QY 189 GAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 248  
Db 142 GAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 201  
QY 249 CATTCCAAATTTCTGCAGTGGACTGCTGGCCAGTGGCATCAAAATTTGATGAATGGGATGCC 308  
Db 202 CATTCCAAATTTCTGCAGTGGACTGCTGGCCAGTGGCATCAAAATTTGATGAATGGGATGCC 261  
QY 309 CACTTTCTCCTGCTCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGCAGTAAGTG 368  
Db 262 CACTTTCTCCTGCTCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGCAGTAAGTG 321  
QY 369 TGGTGAAGGCACCTATTCTTGGCCAGTGGCATCAAAATTTGATGAATGGGATGCC 428  
Db 322 TGGTGAAGGCACCTATTCTTGGCCAGTGGCATCAAAATTTGATGAATGGGATGCC 381  
QY 429 GGCAGGATTTCTAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGACAGCAG 488  
Db 382 GGCAGGATTTCTAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGACAGCAG 441  
QY 489 GCCAGACGGCTGTAAACAACTCTTCTTGGATCCCTCGTGGAAACTACATAGAATCTAATCG 548  
Db 442 GCCAGACGGCTGTAAACAACTCTTCTTGGATCCCTCGTGGAAACTACATAGAATCTAATCG 501  
QY 549 TGATGACTGCACGGTGTCTTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGCTTT 608



Db 502 TGATGACTGCACGGTGTCTTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGTCTT 561  
QY 609 CTTTGAGTACCAAGTATGTCGACAACAACATCTTCTTTGAGTTCCTTTATCAAAATGATCA 668  
Db 562 CTTTGAGTACCAAGTATGTCGACAACAACATCTTCTTTGAGTTCCTTTATCAAAATGATCA 621  
QY 669 GTGCCAGGAGATGGACACCACCACCTGACAAGTGGGTAAAACTTACAGACAATGGAGAATG 728  
Db 622 GTGCCAGGAGATGGACACCACCACCTGACAAGTGGGTAAAACTTACAGACAATGGAGAATG 681  
QY 729 GGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATACTCTACTGGAGAACTACAGG 788  
Db 682 GGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATACTCTACTGGAGAACTACAGG 741  
QY 789 CATCCTTATGGGTTCTAAGCGGGTCAAGCCTGTGCTGGTAAAAAATATCACAAATTAAGG 848.  
Db 742 CATCCTTATGGGTTCTAAGCGGGTCAAGCCTGTGCTGGTAAAAAATATCACAAATTAAGG 801  
QY 849 GGTGGCGTACACATCACAGAAATGTTTTCCTTGCAGCCAGGCACATTCAGCAACAAACCAGG 908  
Db 802 GGTGGCGTACACATCACAGAAATGTTTTCCTTGCAGCCAGGCACATTCAGCAACAAACCAGG 861  
QY 909 TTCATTCAACTGCCAGGTGTGTCGCCAGAAACACCTATTCTGAGAAAGGAGGCCAAAGAA 1082  
Db 862 TTCATTCAACTGCCAGGTGTGTCGCCAGAAACACCTATTCTGAGAAAGGAGGCCAAAGAA 1041  
QY 969 TATAAGGTGTAAAGACGACTCTCAATTTTC-----AGGATCCAGTGAGTGTACAGAGCG 1022  
Db 922 TATAAGGTGTAAAGACGACTCTCAATTTTCAGAGGAAGGATCCAGTGAGTGTACAGAGCG 981  
QY 1023 CCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATACTCCATGTGATGAAGAGGAAA 1082  
Db 982 CCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATACTCCATGTGATGAAGAGGAAA 1041  
QY 1083 GACACAGATAATGTACAAGTGGATAGAGGCCAAAATCTGCCGGGAGGATCTCACAGATGC 1142  
Db 1042 GACACAGATAATGTACAAGTGGATAGAGGCCAAAATCTGCCGGGAGGATCTCACAGATGC 1101  
QY 1143 TATTAGATTGCCCCCTTCTGGAGAGAGAGGATTGTCCGCCCTTGCAACCCCTGGATTTTA 1202  
Db 1102 TATTAGATTGCCCCCTTCTGGAGAGAGAGGATTGTCCGCCCTTGCAACCCCTGGATTTTA 1161  
QY 1203 TAACAATGGATCATCTTCTTGCCCATCCCTGTCCCTCCTGGAACATTTTCAGATGGAACCAA 1262  
Db 1162 TAACAATGGATCATCTTCTTGCCATCCCTGTCCCTCCTGGAACATTTTCAGATGGAACCAA 1221  
QY 1263 AGAATGTAGACCATGTCCAGCAGGAACGGAGCCCTGCACCTTGGCTTTGAATATAAATGGTG 1322  
Db 1222 AGAATGTAGACCATGTCCAGCAGGAACGGAGCCCTGCACCTTGGCTTTGAATATAAATGGTG 1281  
QY 1323 GAATGTCCCTCCTGGCAACATGAAAACCTCCTGCTTCAATGTTGGGAATTCAAAGTGCGA 1382  
Db 1282 GAATGTCCCTCCTGGCAACATGAAAACCTCCTGCTTCAATGTTGGGAATTCAAAGTGCGA 1341  
QY 1383 TGGAAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGA 1442  
Db 1342 TGGAAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGA 1401  
QY 1443 CAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATGAC 1502  
Db 1402 CAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATGAC 1461  
QY 1503 TGGAGCCACGGGTTCTGAACCTAGGARAATAACATTTGTCTTTGAGACCCCTCTGTTTCAGC 1562  
Db 1462 TGGAGCCACGGGTTCTGAACCTAGGARAATAACATTTGTCTTTGAGACCCCTCTGTTTCAGC 1521  
QY 1563 TGACTGTGTTTGTACTTTCATGCTGATTAATAGAAAAAGTACAAATGTGGTAGAATC 1622  
Db 1522 TGACTGTGTTTGTACTTTCATGCTGATTAATAGAAAAAGTACAAATGTGGTAGAATC 1581  
QY 1623 GTGGGGTGAACCAAGAAAAACAGCTTACACCCCATATCATCTTCAAGAAATGCAACTTT 1682  
|||||

Db 1582 GTGGGGTGGAAACCAAGAAAAACAAGCTTACACCCATATCATCTTCAAGAANTGCAACTTTT 1641  
QY 1683 TACATTTACATGGG--CATTCAGAGAACTAANT--CAGGGTCAAGATAAATAGACGGTTCA 1738  
Db 1642 TACATTTACATGGGGCATTTCCCAAGAGAACTAANTTCAGGGTCCCAAGATAAATAGACGGTTCC 1701  
QY 1739 TCAAT 1743  
Db 1702 NCCAT 1706  
RESULT 4  
US-10-002-304-19  
; Sequence 19, Application US/10002304  
; Publication No. US20030036185A1  
; GENERAL INFORMATION:  
; APPLICANT: Shinkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Hermann, John  
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby  
; FILE REFERENCE: 15966-554 Cura-54 CON-S8  
; CURRENT APPLICATION NUMBER: US/10/002,304  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 1737  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (296)..(1687)  
; NAME/KEY: variation  
; LOCATION: (1)..(1737)  
; OTHER INFORMATION: N may be any nucleotide  
US-10-002-304-19

Query Match 42.2%; Score 1499.6; DB 9; Length 1737;  
Best Local Similarity 98.7%; pred. No. 0;  
Matches 1545; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

QY 189 GAAAGATTATCAGCTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 248  
Db 142 GAAAGATTATCAGCTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 201  
QY 249 CATTCCAAATTTCTGCAGTGGAGTGTGCTGGCCTGCCTGACCCAGTGAGAGGCAAGAATG 308  
Db 202 CATTCCAAATTTCTGCAGTGGAGTGTGCTGGCCTGCCTGACCCAGTGAGAGGCAAGAATG 261  
QY 309 CACTTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCAGGTATGCAGTAAGTG 368  
Db 262 CACTTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCAGGTATGCAGTAAGTG 321  
QY 369 TGGTGAAGGCACCTATTCCCTTGGGCAGTGGCATCAAAATTTGATGAATGGGATGAATGCC 428  
Db 322 TGGTGAAGGCACCTATTCCCTTGGGCAGTGGCATCAAAATTTGATGAATGGGATGAATGCC 381  
QY 429 GGCAGGATTTTCTAACATCCCAACATTCATGGACACTGTGGTGGGCCCTTCTGACAGCAG 488  
Db 382 GGCAGGATTTTCTAACATCCCAACATTCATGGACACTGTGGTGGGCCCTTCTGACAGCAG 441  
QY 489 GCCAGACGGCTGTAACTACTTCTTGGATCCCTCGTGGAAACTACATAGAAATCTAATCG 548  
Db 442 GCCAGACGGCTGTAACTACTTCTTGGATCCCTCGTGGAAACTACATAGAAATCTAATCG 501  
QY 549 TGATGACTGCACGGTGTCTTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATCTCTT 608

Db 502 TGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGTCTT 561  
Qy 609 CTTTGAGTACCAGTATGTCGACACAACATCTTCTTTGAGTTCCTTATTCAAAATGATCA 668  
Db 562 CTTTGAGTACCAGTATGTCGACACAACATCTTCTTTGAGTTCCTTATTCAAAATGATCA 621  
Qy 669 GTGCCAGGAGATGGACACCACCTGACAAGTGGGTAAACCTTACAGACAATGGAGAATG 728  
Db 622 GTGCCAGGAGATGGACACCACCTGACAAGTGGGTAAACCTTACAGACAATGGAGAATG 681  
Qy 729 GGGCTCTCATTTCTGAATGCTGAATCAGGCACAAAACATACCTACTCTGAGAACTACAGG 788  
Db 682 GGGCTCTCATTTCTGAATGCTGAATCAGGCACAAAACATACCTACTCTGAGAACTACAGG 741  
Qy 789 CATCCTATATGGGTTCTAAGCGGTCAAGCCTGTGCTGGTAAAAAATATCACAATTTGAAGG 848  
Db 742 CATCCTATATGGGTTCTAAGCGGTCAAGCCTGTGCTGGTAAAAAATATCACAATTTGAAGG 801  
Qy 849 GGTGGCGTACACATCAGAATGTTTCCTTGCAAGCCAGGCACATTCAGCAACAACCCAGG 908  
Db 802 GGTGGCGTACACATCAGAATGTTTCCTTGCAAGCCAGGCACATTCAGCAACAACCCAGG 861  
Qy 909 TTCATTCAACTGCCAGGTGTGCCAGAAAACACCTATTCTGAAAAGGAGCCAAAAGAATG 968  
Db 862 TTCATTCAACTGCCAGGTGTGCCAGAAAACACCTATTCTGAAAAGGAGCCAAAAGAATG 921  
Qy 969 TATAAGGTGTAAAGACGACTCTCAATTTTC-----AGGATCCAGTGAGTGTACAGAGCG 1022  
Db 922 TATAAGGTGTAAAGACGACTCTCAATTTTCAGAGGAAGGATCCAGTGAGTGTACAGAGCG 981  
Qy 1023 CCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATACTCCATGTGATGAAGAAGGAAA 1082  
Db 982 CCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATACTCCATGTGATGAAGAAGGAAA 1041  
Qy 1083 GACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGAGGATCTCACAGATGC 1142  
Db 1042 GACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGAGGATCTCACAGATGC 1101  
Qy 1143 TATTAGATTGCCCCCTTCTGGAGAGAAGAGGATTGTCCGCCCTTGCAACCCCTGGATTTTA 1202  
Db 1102 TATTAGATTGCCCCCTTCTGGAGAGAAGAGGATTGTCCGCCCTTGCAACCCCTGGATTTTA 1161  
Qy 1203 TAACAAATGGATCATCTTCTTGCCATCCCTGTCCCTGGAACATTTTCAGATGGAACCAA 1262  
Db 1162 TAACAAATGGATCATCTTCTTGCCATCCCTGTCCCTGGAACATTTTCAGATGGAACCAA 1221  
Qy 1263 AGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAATGGTG 1322  
Db 1222 AGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAATGGTG 1281  
Qy 1323 GAATGTCCTTCTGGCAACATGAAAACCTTCTGCTTCAATGTTGGGAATTCAAAAGTGCGA 1382  
Db 1282 GAATGTCCTTCTGGCAACATGAAAACCTTCTGCTTCAATGTTGGGAATTCAAAAGTGCGA 1341  
Qy 1383 TGGAAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGA 1442  
Db 1342 TGGAAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGA 1401  
Qy 1443 CAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCAACCAACATCTATGAC 1502  
Db 1402 CAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCAACCAACATCTATGAC 1461  
Qy 1503 TGGAGCCACGGGTTCTGAACCTAGGAAGAAATAACATTTGCTTTTGAGACCCCTCTGTTTCAGC 1562  
Db 1462 TGGAGCCACGGGTTCTGAACCTAGGAAGAAATAACATTTGCTTTTGAGACCCCTCTGTTTCAGC 1521  
Qy 1563 TGACTGTGTTTGTACTTTCATGGTGGATATTAATAGAAAAAGTACAAAATGTGGTAGAATC 1622  
Db 1522 TGACTGTGTTTGTACTTTCATGGTGGATATTAATAGAAAAAGTACAAAATGTGGTAGAATC 1581  
Qy 1623 GTGGGGTGAACCAAGAAAAACACCTTACACCCATATCATCTTCAAGAATGCAACTTT 1682

Db 1582 GTGGGGTGAACCAAGAAAAACAAGCTTACACCCATATCATCTTCAAGAATGCAACTTT 1641  
Qy 1683 TACATTTACATGGG--CATTCAGAGAACTAAT--CAGGGTCAAGATAATAGACGGTTCA 1738  
Db 1642 TACATTTACATGGGCATTTCCAGAGAACTAATTCAGGGTCCAAGATAATAGACGGTTCC 1701  
Qy 1739 TCAAT 1743  
Db 1702 NCCAT 1706  
RESULT 5  
US-10-003-152-19  
; Sequence 19, Application US/10003152  
; Patent No. US20020151494A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: No. US20020151494A1el Amino Acid Sequences for Human Semaphori  
; FILE REFERENCE: 15966-554 Cura-54 CON-S12  
; CURRENT APPLICATION NUMBER: US/10/003,152  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 1737  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (296)..(1687)  
; NAME/KEY: variation  
; LOCATION: (1)..(1737)  
; OTHER INFORMATION: N may be any nucleotide  
US-10-003-152-19

Query Match 42.2%; Score 1499.6; DB 12; Length 1737;  
Best Local Similarity 98.7%; Pred. No. 0;  
Matches 1545; Conservative 0; Mismatches 10; Indels 10; Gaps 3;  
Qy 189 GAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCAGGTGGAGAGTTGC 248  
Db 142 GAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 201  
Qy 249 CATTCCAAATTCGCAGTGGACTGCTGGCCTGCCTGACCCAGTGAGAGGCAAAGAATG 308  
Db 202 CATTCCAAATTCGCAGTGGACTGCTGGCCTGCCTGACCCAGTGAGAGGCAAAGAATG 261  
Qy 309 CACTTTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCAGGTATGCAGTAAGTG 368  
Db 262 CACTTTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCAGGTATGCAGTAAGTG 321  
Qy 369 TGGTGAAGGCACCTATTCCCTTGGGCAGTGGCATCAAATTTGATGAATGGGATGAATTGCC 428  
Db 322 TGGTGAAGGCACCTATTCCCTTGGGCAGTGGCATCAAATTTGATGAATGGGATGAATTGCC 381  
Qy 429 GGCAGGATTTCTAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGACAGCAG 488  
Db 382 GGCAGGATTTCTAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGACAGCAG 441  
Qy 489 GCCAGACGGCTGTAACAACTCTTCTGGATCCCTCGTGGAAACTACATAGAATCTAATCG 548  
Db 442 GCCAGACGGCTGTAACAACTCTTCTTGGATCCCTCGTGGAAACTACATAGAATCTAATCG 501  
Qy 549 TGATGACTGCACGGGTGCTTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGCTTT 608

Db 502 TGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAAGAAAGTCAGGCTATGCTTT 561  
Qy 609 CTTTGACTACCAGTATGTCGACAAACATCTTCTTTGAGTCTTTATTCAAAATGATCA 668  
Db 562 CTTTGAGTACCAGTATGTCGACAAACATCTTCTTTGAGTCTTTATTCAAAATGATCA 621  
Qy 669 GTGCCAGGAGATGGACACCACCTGACAAAGTGGGTAAACCTTACAGACAATGGAGAATG 728  
Db 622 GTGCCAGGAGATGGACACCACCTGACAAAGTGGGTAAACCTTACAGACAATGGAGAATG 681  
Qy 729 GGGCTCTCATTTCTAATGCTGAATCAGGCACAAACATACTCTACTGGAGAACTACAGG 788  
Db 682 GGGCTCTCATTTCTAATGCTGAATCAGGCACAAACATACTCTACTGGAGAACTACAGG 741  
Qy 789 CATCCTTATGGGTTCTAAGGGCGTCAAGCCCTGTGCTGGTAAAAAATATCACAATTTGAAG 848  
Db 742 CATCCTTATGGGTTCTAAGGGCGTCAAGCCCTGTGCTGGTAAAAAATATCACAATTTGAAG 801  
Qy 849 GGTGGCSTACACATCAGAAATGTTTCCCTTGCAAGCCAGGCACATTCAGCAACAAACCAGG 908  
Db 802 GGTGGCSTACACATCAGAAATGTTTCCCTTGCAAGCCAGGCACATTCAGCAACAAACCAGG 861  
Qy 909 TTCAATCAACTGCCAGGTGTGCCAGAAACACCTATTCTGAAAGAGGAGCCAAAGAATG 968  
Db 862 TTCAATCAACTGCCAGGTGTGCCAGAAACACCTATTCTGAAAGAGGAGCCAAAGAATG 921  
Qy 969 TATAAGGTGTAAAGACGACTCTCAATTTTC-----AGGATCCAGTGAGTGTACAGAGCG 1022  
Db 922 TATAAGGTGTAAAGACGACTCTCAATTTTCAGAGGAAGGATCCAGTGAGTGTACAGAGCG 981  
Qy 1023 CCCTCCCTGTACCACAAAGACTATTTCAGAGATCCATCTCCATGTGATGAAGAGGAAA 1082  
Db 982 CCCTCCCTGTACCACAAAGACTATTTCAGAGATCCATCTCCATGTGATGAAGAGGAAA 1041  
Qy 1083 GACACAGATAATGTACAAAGTGGATAGAGCCCCAAAATCTGCCGGAGGATCTCACAGATGC 1142  
Db 1042 GACACAGATAATGTACAAAGTGGATAGAGCCCCAAAATCTGCCGGAGGATCTCACAGATGC 1101  
Qy 1143 TATTAGATTGCCCCCTTCTGGAGAGAAGGATTGTCCGCCCTTGCAACCCCTGGATTTTA 1202  
Db 1102 TATTAGATTGCCCCCTTCTGGAGAGAAGGATTGTCCGCCCTTGCAACCCCTGGATTTTA 1161  
Qy 1203 TAACAAATGGATCATCTTCTTGCCATCCCTGTCTCCTCGAACATTTTCAGATGGAACCAA 1262  
Db 1162 TAACAAATGGATCATCTTCTTGCCATCCCTGTCTCCTCGAACATTTTCAGATGGAACCAA 1221  
Qy 1263 AGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGACACTTGGCTTTGAATATAAATGGTG 1322  
Db 1222 AGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGACACTTGGCTTTGAATATAAATGGTG 1281  
Qy 1323 GAATGTCTTCTCTGGCAACATGAAACTTCTCTGCTTCAATGTTGGGAATTCAAAGTGCGA 1382  
Db 1282 GAATGTCTTCTCTGGCAACATGAAACTTCTCTGCTTCAATGTTGGGAATTCAAAGTGCGA 1341  
Qy 1383 TGGAAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGA 1442  
Db 1342 TGGAAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGA 1401  
Qy 1443 CAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATGAC 1502  
Db 1402 CAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATGAC 1461  
Qy 1503 TGGAGCCACGGGTTCTGAACTAGGAAGAATAACATTTGTCTTTGAGACCCCTCTGTTTCAGC 1562  
Db 1462 TGGAGCCACGGGTTCTGAACTAGGAAGAATAACATTTGTCTTTGAGACCCCTCTGTTTCAGC 1521  
Qy 1563 TGACTGTCTTTTGTACTTTCATGGTGGATATTAATAGAAAAAGTACAAAATGTGGTAGAATC 1622  
Db 1522 TGACTGTGTGTTGTACTTTCATGGTGGATATTAATAGAAAAAGTACAAAATGTGGTAGAATC 1581  
Qy 1623 GTGGGGTGGAAACCAAGAAAAACACCTTACACCCCATATCATCTTCAAGAAATGCAACTTT 1682

Db 1582 GTGGGTGGAACCAAGAAAAACAAAGCTTACACCCATATCATCTTCAAGAATGCAACTTT 1641  
Qy 1683 TACATTTACATGGG--CATTCAGAGAACTAAT--CAGGGTCAAGATAATAGACGGTTCA 1738  
Db 1642 TACATTTACATGGGCATTTCCAGAGAACTAATTCAGGGTCCAAGATAATAGACGGTTCC 1701  
Qy 1739 TCAAT 1743  
Db 1702 NCCAT 1706  
RESULT 6  
US-10-002-050-9  
; Sequence 9, Application US/10002050  
; Publication No. US20030032095A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meljia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: No. US20030032095A1el Nucleic Acid Sequences Encoding Human Se  
; FILE REFERENCE: 15966-554 Cura-54 CON-S14  
; CURRENT APPLICATION NUMBER: US/10/002,050  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604,286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140,584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 1508  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (226)..(1458)  
; NAME/KEY: variation  
; LOCATION: (1)..(1508)  
; OTHER INFORMATION: N may be any nucleotide  
US-10-002-050-9

Query Match 33.3%; Score 1182.6; DB 9; Length 1508;

Best Local Similarity 98.4%; Pred. No. 0;

Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

Qy 506 ACTCTTCTTGGATCCCTCGTGGAACTACATAGAATCTAATCGTGATGACTGCACGGTGT 565  
Db 230 ACTCTTCTTGGATCCCTCGTGGAACTACATAGAATCTAATCGTGATGACTGCACGGTGT 289  
Qy 566 CTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGTCTCTTTGAGTACCAGTATG 625  
Db 290 CTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGTCTCTTTGAGTACCAGTATG 349  
Qy 626 TCGACAACAACATCTTCTTTGAGTCTTTATTCAAAATGATCAGTGCACGAGAGATGGACA 685  
Db 350 TCGACAACAACATCTTCTTTGAGTCTTTATTCAAAATGATCAGTGCACGAGAGATGGACA 409  
Qy 686 CCACCACCTGACAAGTGGGTAAACCTTACAGACAATGGAGAATGGGCTCTCATTCGTAA 745  
Db 410 CCACCACCTGACAAGTGGGTAAACCTTACAGACAATGGAGAATGGGCTCTCATTCGTAA 469  
Qy 746 TGCTGAAATCAGGCACAAACATACTCTACTGGAGAACTACAGGCACTCCTTATGGGTCTA 805  
Db 470 TGCTGAAATCAGGCACAAACATACTCTACTGGAGAACTACAGGCACTCCTTATGGGTCTA 529  
Qy 806 AGCGGTCAAAGCCTGTGCTGGTAAAAAATATCACAAATTTGAAGGGTGGCGGTACACATCAG 865  
Db 530 AGCGGTCAAAGCCTGTGCTGGTAAAAAATATCACAAATTTGAAGGGTGGCGGTACACATCAG 589  
Qy 866 AATGTTTTCCCTTGCAAGCCAGGCACATTCAGCAACAAACCAGGTTTCATTCAAACTGCCCAGG 925



|||||  
Db 590 AATGTTTCTTGAAGCCAGGCACATTTCAGCAACAACCCAGGTTTCATTCAACTGCCAGG 649  
QY 926 TGTGTCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAATGTATAAGGTGTAAAGACG 985  
Db 650 TGTGTCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAATGTATAAGGTGTAAAGACG 709  
QY 986 ACTCTCAATTTTC-----AGGATCCAGTGAGTGTACAGAGCGCCCTCCCTGTACCACAA 1039  
Db 710 ACTCTCAATTTTCAGAGGAAGGATCCAGTGAGTGTACAGAGCGCCCTCCCTGTACCACAA 769  
QY 1040 AAGACTATTTCAGATCCATCTCCATGTGATGAAGAAAGGAAAGACACAGATAATGTACA 1099  
Db 770 AAGACTATTTCAGATCCATCTCCATGTGATGAAGAAAGGAAAGACACAGATAATGTACA 829  
QY 1100 AGTGATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGATGCTATTAGATTGCCCCCTT 1159  
Db 830 AGTGATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGATGCTATTAGATTGCCCCCTT 889  
QY 1160 CTGGAGAGAAGAGGATTGTCCGCCCTTGAACCCCTGGATTGTATAACAATGGATCATCTT 1219  
Db 890 CTGGAGAGAAGAGGATTGTCCGCCCTTGAACCCCTGGATTGTATAACAATGGATCATCTT 949  
QY 1220 CTTGCCATCCCTGTCTCCTCGAACATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1279  
Db 950 CTTGCCATCCCTGTCTCCTCGAACATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1009  
QY 1280 CAGCAGGAACGGAGCCCTGCACTTGGCTTTGAATATAAATGGTGGAAATGTCCTTCCTGGCA 1339  
Db 1010 CAGCAGGAACGGAGCCCTGCACTTGGCTTTGAATATAAATGGTGGAAATGTCCTTCCTGGCA 1069  
QY 1340 ACATGAAAACCTCCTGCTTCAATGTGGGAATTCAAAGTGGATGAATGAATGGTTGGG 1399  
Db 1070 ACATGAAAACCTCCTGCTTCAATGTGGGAATTCAAAGTGGATGAATGAATGGTTGGG 1129  
QY 1400 AGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGACAATGATTACCTGATCT 1459  
Db 1130 AGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGACAATGATTACCTGATCT 1189  
QY 1460 TAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGGTTCTG 1519  
Db 1190 TAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGGTTCTG 1249  
QY 1520 AACTAGGAAGAATAACATTTGCTTTTGAGACCCCTCTGTTTCAGCTGACTGTGTTTGTACT 1579  
Db 1250 AACTAGGAAGAATAACATTTGCTTTTGAGACCCCTCTGTTTCAGCTGACTGTGTTTGTACT 1309  
QY 1580 TCATGGTGGATATTAATAGAAAAGTACAAAATGTGGTAGAATCGTGGGGTGGAAACCAAG 1639  
Db 1310 TCATGGTGGATATTAATAGAAAAGTACAAAATGTGGTAGAATCGTGGGGTGGAAACCAAG 1369  
QY 1640 AAAACAAGCTTACACCCATATCATCTTCAAGAAATGCAACTTTTACATTTACATGGG--C 1697  
Db 1370 AAAACAAGCTTACACCCATATCATCTTCAAGAAATGCAACTTTTACATTTACATGGGGCA 1429  
QY 1698 ATTCAGAGAACTAAT--CAGGGTCAAGATAAATAGACGGTTTCATCAAT 1743  
Db 1430 TTCCCAGAGAACTAATTCAGGGTCCAAGATAAATAGACGGTTTCNCCAT 1477

RESULT 7  
US-10-002-304-9  
; Sequence 9, Application US/10002304  
; Publication No. US20030036185A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Vernet, Corine  
; APPLICANT: Yang, Meijia  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Herrmann, John  
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby  
; FILE REFERENCE: 15966-554 Cura-54 CON-S8

; CURRENT APPLICATION NUMBER: US/10/002.304  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604.286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140.584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 1508  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (226)..(1458)  
; NAME/KEY: variation  
; LOCATION: (1)..(1508)  
; OTHER INFORMATION: N may be any nucleotide  
US-10-002-304-9

Query Match 33.3%; Score 1182.6; DB 9; Length 1508;  
Best Local Similarity 98.4%; Pred. No. 0;  
Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;  
QY 506 ACTCTTCTTGGATCCCTCGTGGAAACTACATAGAAATCTAATCGTGATGACTGCACGGTGT 565  
Db 230 ACTCTTCTTGGATCCCTCGTGGAAACTACATAGAAATCTAATCGTGATGACTGCACGGTGT 289  
QY 566 CTTTGATCTATGCTGTGCACCTTAAAGAGTCAGGCTATGTCTTCTTTGAGTACCAGTATG 625  
Db 290 CTTTGATCTATGCTGTGCACCTTAAAGAGTCAGGCTATGTCTTCTTTGAGTACCAGTATG 349  
QY 626 TCGACACAACAACATCTTCTTTGAGTCTTTATTCAAATGATCATGTCGCCAGGAGATGGACA 685  
Db 350 TCGACACAACAACATCTTCTTTGAGTCTTTATTCAAATGATCATGTCGCCAGGAGATGGACA 409  
QY 686 CCACCACCTGACAAGTGGGTAAAACTTACAGACAATGGAGAAATGGGGCTCTCATTTCTGTAA 745  
Db 410 CCACCACCTGACAAGTGGGTAAAACTTACAGACAATGGAGAAATGGGGCTCTCATTTCTGTAA 469  
QY 746 TGCTGAAATCAGGCACAAACATACCTACTGGAGAACTACAGGCATCCTTATGGGTTCTA 805  
Db 470 TGCTGAAATCAGGCACAAACATACCTACTGGAGAACTACAGGCATCCTTATGGGTTCTA 529  
QY 806 AGCGGTCAAGCCTGTGCTGGTAAAAAATATCACAATTTGAAGGGTGGCGTACACATCAG 865  
Db 530 AGCGGTCAAGCCTGTGCTGGTAAAAAATATCACAATTTGAAGGGTGGCGTACACATCAG 589  
QY 866 AATGTTTTCCTTGAAGCCAGGCACATTCAGCAACAACCCAGGTTTCATTCAACTGCCAGG 925  
Db 590 AATGTTTTCCTTGAAGCCAGGCACATTCAGCAACAACCCAGGTTTCATTCAACTGCCAGG 649  
QY 926 TGTGTCCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAATGTATAAGGTGTAAAGACG 985  
Db 650 TGTGTCCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAATGTATAAGGTGTAAAGACG 709  
QY 986 ACTCTCAATTTTC-----AGGATCCAGTGAGTGTACAGAGCGCCCTCCCTGTACCACAA 1039  
Db 710 ACTCTCAATTTTCAGAGGAAGGATCCAGTGAGTGTACAGAGCGCCCTCCCTGTACCACAA 769  
QY 1040 AAGACTATTTCAGATCCATCTCCATGTGATGAAGAAAGGAAAGACACAGATAATGTACA 1099  
Db 770 AAGACTATTTCAGATCCATCTCCATGTGATGAAGAAAGGAAAGACACAGATAATGTACA 829  
QY 1100 AGTGATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGATGCTATTAGATTGCCCCCTT 1159  
Db 830 AGTGATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGATGCTATTAGATTGCCCCCTT 889  
QY 1160 CTGGAGAGAAGAGGATTGTCCGCCCTTCCAACCCCTGGATTTTATACAATGGATCATCTT 1219  
Db 890 CTGGAGAGAAGAGGATTGTCCGCCCTTCCAACCCCTGGATTTTATACAATGGATCATCTT 949  
QY 1220 CTTGCCATCCCTGTCTCCTCTCGGAACATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1279

```

Db 950 CTTGCCATCCCTGTCCCTCGGAACATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1009
Qy 1280 CAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAAATGGTGAATGTCCCTCGGCA 1339
Db 1010 CAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAAATGGTGAATGTCCCTCGGCA 1069
Qy 1340 ACATGAAACTTCCCTGCTTCAATGTTGGGAATTCAAAGTGGGATGGAATGAATGGTTGGG 1399
Db 1070 ACATGAAACTTCCCTGCTTCAATGTTGGGAATTCAAAGTGGGATGGAATGAATGGTTGGG 1129
Qy 1400 AGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGACAATGATTACCTGATCT 1459
Db 1130 AGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGACAATGATTACCTGATCT 1189
Qy 1460 TAAACTTGCATATCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGGTTCTG 1519
Db 1190 TAAACTTGCATATCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGGTTCTG 1249
Qy 1520 AACTAGGAAGAAATAACATTTGCTTTTGAGACCCCTCTGTTCCAGCTGACTGTGTTTGACT 1579
Db 1250 AACTAGGAAGAAATAACATTTGCTTTTGAGACCCCTCTGTTCCAGCTGACTGTGTTTGACT 1309
Qy 1580 TCATGGTGGATTAATAGAAAAAGTACAAAATGTGGTAGAATCGTGGGGTGGAACCAAG 1639
Db 1310 TCATGGTGGATTAATAGAAAAAGTACAAAATGTGGTAGAATCGTGGGGTGGAACCAAG 1369
Qy 1640 AAAACAACGCTTACACCCATATCATCTTCAAGAATGCAACTTTTACATTTACATGGG--C 1697
Db 1370 AAAACAACGCTTACACCCATATCATCTTCAAGAATGCAACTTTTACATTTACATGGGCA 1429
Qy 1698 ATTCCAGAACTAAT--CAGGGTCAAGATAAATAGACGGTTTCATCAAT 1743
Db 1430 TTCCCAGAGAACTAATTCAGGGTCCAAGATAATAGACGGTTCCNCCAT 1477

RESULT 8
US-10-003-152-9
; Sequence 9, Application US/10003152
; Patent No. US20020151494A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meljia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20020151494A1el Amino Acid Sequences for Human Semaphorin-1
; FILE REFERENCE: 15966-554 Cura-54 CON-S12
; CURRENT APPLICATION NUMBER: US/10/003,152
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1508
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (226)..(1458)
; NAME/KEY: variation
; LOCATION: (1)..(1508)
; OTHER INFORMATION: N may be any nucleotide
US-10-003-152-9
Query Match 33.3%; Score 1182.6; DB 12; Length 1508;
Best Local Similarity 98.4%; Pred. No.: 0;
Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;
```

```

Qy 506 ACTCTTCTTGGATCCCTCGTGGAAACTACATAGAAATCTAATCGTGATGACTGCACGGTGT 565
Db 230 ACTCTTCTTGGATCCCTCGTGGAAACTACATAGAAATCTAATCGTGATGACTGCACGGTGT 289
Qy 566 CTTTGATCTATGCTGTGCACCTTAAGAGTCAAGGCTATGTCTTCTTTAGTACCAGTATG 625
Db 290 CTTTGATCTATGCTGTGCACCTTAAGAGTCAAGGCTATGTCTTCTTTAGTACCAGTATG 349
Qy 626 TCGACAACAACATCTTCTTTGAGTCTTTATTCMAAATGATCAGTGCAGGAGATGGACA 685
Db 350 TCGACAACAACATCTTCTTTGAGTCTTTATTCMAAATGATCAGTGCAGGAGATGGACA 409
Qy 686 CCACCACCTGACAAGTGGGTAAAACITACAGACAATGGAGAATGGGGCTCTCATTTCTATA 745
Db 410 CCACCACCTGACAAGTGGGTAAAACITACAGACAATGGAGAATGGGGCTCTCATTTCTATA 469
Qy 746 TGCTGAATCAGGCACAAACATACTCTACTGGAGAACTACAGGCATCCTTATGGGTTCTA 805
Db 470 TGCTGAATCAGGCACAAACATACTCTACTGGAGAACTACAGGCATCCTTATGGGTTCTA 529
Qy 806 AGCGGTCAAGCCTGTGCTGTTAAATAATTCACAAATGAAGGGTGGCGTACACATCAG 865
Db 530 AGCGGTCAAGCCTGTGCTGTTAAATAATTCACAAATGAAGGGTGGCGTACACATCAG 589
Qy 866 AATGTTTTCCTTGCAAGCCAGGCACATTCAGCAACAAACCCAGGTTTCATTCAACTGCCAGG 925
Db 590 AATGTTTTCCTTGCAAGCCAGGCACATTCAGCAACAAACCCAGGTTTCATTCAACTGCCAGG 649
Qy 926 TGTGTCCAGAAACACCTATCTGAGAAAGGAGCCAAAGAATGTATAAGGTGTAAAGACG 985
Db 650 TGTGTCCAGAAACACCTATCTGAGAAAGGAGCCAAAGAATGTATAAGGTGTAAAGACG 709
Qy 986 ACTCTCAATTTTC-----AGGATCCAGTGAAGTGTACAGAGCGGCCCTCCCTGTACCACAA 1039
Db 710 ACTCTCAATTTTCAGAGGAAGGATCCAGTGAAGTGTACAGAGCGGCCCTCCCTGTACCACAA 769
Qy 1040 AAGACTATTTCCAGATCCATACTCCCTGTGATGAAGAGGAAGAACACACAGATAATGTACA 1099
Db 770 AAGACTATTTCCAGATCCATACTCCCTGTGATGAAGAGGAAGAACACACAGATAATGTACA 829
Qy 1100 AGTGGATAGAGCCCCAAATCTGCCGGGAGGATCTCACAGATGCTATTAGATTGCCCCCTT 1159
Db 830 AGTGGATAGAGCCCCAAATCTGCCGGGAGGATCTCACAGATGCTATTAGATTGCCCCCTT 889
Qy 1160 CTGGAGAGAAGAAGGATGTCCGCCCTTGCAACCCCTGGATTTTATAACAATGGATCATCTT 1219
Db 890 CTGGAGAGAAGAAGGATGTCCGCCCTTGCAACCCCTGGATTTTATAACAATGGATCATCTT 949
Qy 1220 CTTGCCATCCCTGTCCCTCGTGAACATTTTCAGATGGAACCAACAAAGATGTAGACCATGTC 1279
Db 950 CTTGCCATCCCTGTCCCTCGTGAACATTTTCAGATGGAACCAACAAAGATGTAGACCATGTC 1009
Qy 1280 CAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAAATGGTGAATGTCCCTCGGCA 1339
Db 1010 CAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAAATGGTGAATGTCCCTCGGCA 1069
Qy 1340 ACATGAAACTTCCCTGCTTCAATGTTGGGAATTCAAAGTGGGATGGAATGAATGGTTGGG 1399
Db 1070 ACATGAAACTTCCCTGCTTCAATGTTGGGAATTCAAAGTGGGATGGAATGAATGGTTGGG 1129
Qy 1400 AGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGACAATGATTACCTGATCT 1459
Db 1130 AGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGACAATGATTACCTGATCT 1189
Qy 1460 TAAACTTGCATATCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGGTTCTG 1519
Db 1190 TAAACTTGCATATCCAGGATTTAAACCACCAACATCTATGACTGGAGCCACGGGTTCTG 1249
Qy 1520 AACTAGGAAGAAATAACATTTGCTTTTGAGACCCCTCTGTTCCAGCTGACTGTGTTTGACT 1579
Db 1250 AACTAGGAAGAAATAACATTTGCTTTTGAGACCCCTCTGTTCCAGCTGACTGTGTTTGACT 1309
Qy 1580 TCATGGTGGATTAATAGAAAAAGTACAAAATGTGGTAGAATCGTGGGGTGGAACCAAG 1639
```

Db 1310 TCATGGTGGATATAATAGAAAAGTACAATGTGGTAGAATCGTGGGTGGAACCAAG 1369  
QY 1640 AAAACAAGCTTACACCCATATCATCTTCAAGAAATGCAACTTTTACATTTACATGGG--C 1697  
Db 1370 AAAACAAGCTTACACCCATATCATCTTCAAGAAATGCAACTTTTACATTTACATGGGCA 1429  
QY 1698 ATTCCAGAGAACTAAT--CAGGGTCAAGATAATAGACGGTTTCATCAAT 1743  
Db 1430 TTCCCAGAGAACTAATTCAGGGTCCAAGATAATAGACGGTTCNCNCCAT 1477

RESULT 9

US-10-028-072-37  
; Sequence 37, Application US/10028072  
; Publication No. US20030004311A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang  
; TITLE OF INVENTION:  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/028,072  
; CURRENT FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/049911  
; PRIOR FILING DATE: 1997-06-18  
; PRIOR APPLICATION NUMBER: 60/056974  
; PRIOR FILING DATE: 1997-08-26  
; PRIOR APPLICATION NUMBER: 60/059113  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059115  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059117  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059122  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059184  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059352  
; PRIOR FILING DATE: 1997-09-19  
; PRIOR APPLICATION NUMBER: 60/059588  
; PRIOR FILING DATE: 1997-09-19  
; PRIOR APPLICATION NUMBER: 60/059836  
; PRIOR FILING DATE: 1997-09-24  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/062285  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/062287  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/062814  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/062816  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063045  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063082

; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: 60/063127  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063327  
; PRIOR FILING DATE: 1997-10-27  
; PRIOR APPLICATION NUMBER: 60/063329  
; PRIOR FILING DATE: 1997-10-27  
; PRIOR APPLICATION NUMBER: 60/063550  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063561  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063704  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/063733  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/063735  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/063738  
; PRIOR FILING DATE: 1997-10-29  
; PRIOR APPLICATION NUMBER: 60/063755  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/064248  
; PRIOR FILING DATE: 1997-11-03  
; PRIOR APPLICATION NUMBER: 60/064809  
; PRIOR FILING DATE: 1997-11-07  
; PRIOR APPLICATION NUMBER: 60/065186  
; PRIOR FILING DATE: 1997-11-12  
; PRIOR APPLICATION NUMBER: 60/065846  
; PRIOR FILING DATE: 1997-11-17  
; PRIOR APPLICATION NUMBER: 60/066364  
; PRIOR FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: 60/066453  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/066511  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/066770  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/069212  
; PRIOR FILING DATE: 1997-12-11  
; PRIOR APPLICATION NUMBER: 60/069278  
; PRIOR FILING DATE: 1997-12-11  
; PRIOR APPLICATION NUMBER: 60/069334  
; PRIOR FILING DATE: 1997-12-11  
; PRIOR APPLICATION NUMBER: 60/069694  
; PRIOR FILING DATE: 1997-12-16  
; PRIOR APPLICATION NUMBER: 60/072320  
; PRIOR FILING DATE: 1998-01-23  
; PRIOR APPLICATION NUMBER: 60/073612  
; PRIOR FILING DATE: 1998-02-04  
; PRIOR APPLICATION NUMBER: 60/074086  
; PRIOR FILING DATE: 1998-02-09  
; PRIOR APPLICATION NUMBER: 60/074092  
; PRIOR FILING DATE: 1998-02-09  
; PRIOR APPLICATION NUMBER: 60/077791  
; PRIOR FILING DATE: 1998-03-12  
; PRIOR APPLICATION NUMBER: 60/078910  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/079294  
; PRIOR FILING DATE: 1998-03-25  
; PRIOR APPLICATION NUMBER: 60/079663  
; PRIOR FILING DATE: 1998-02-27  
; PRIOR APPLICATION NUMBER: 60/079728  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: 60/080165  
; PRIOR FILING DATE: 1998-03-31  
; PRIOR APPLICATION NUMBER: 60/081203  
; PRIOR FILING DATE: 1998-04-09  
; PRIOR APPLICATION NUMBER: 60/081229  
; PRIOR FILING DATE: 1998-04-09  
; PRIOR APPLICATION NUMBER: 60/081695  
; PRIOR FILING DATE: 1998-04-14  
; PRIOR APPLICATION NUMBER: 60/081817  
; PRIOR FILING DATE: 1998-04-15





Db 1338 CTCCATGG---CTCAGACTGTACCCGCTGCCCTGCAGGACTGAACCTGCTGTGGGATT 1394  
QY 1308 TGAATATAAATGGTGGAAATGCTTCCCTGGCAACATGAAAACTTCCTGCTTCAATGTTGG 1367  
Db 1395 TGAATACAAATGGTGGAAACACGCTGCCACAAACATGGAACGACCGTTCTCAGTGGGAT 1454  
QY 1368 GAATTCAAAGTGGATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427  
Db 1455 CAACCTCGAGTACAAGGGCATGACAGGCTGGGAGGTGGCTGGTGTATCACATTTACACAGC 1514  
QY 1428 GGCTGGAGGTTCTGACAATGATTACCTGATCTTTAACTTGCATATCCCAGGATTTAAACC 1487  
Db 1515 TGCTGGAGCCTCAGACAATGACTTTCATGATTCTCACTCTGGTTGTGCCAGGATTTAGACC 1574  
QY 1488 ACCAACATCTATGACTGGAGCCAC--GGGTTCTGAACTAGGAAGAATAACATTTGTCTT 1544  
Db 1575 TCCGCAGTCGGTGATGGCAGACACAGAGAATAAAGAGGTGGCCAGAAATCACATTTGTCTT 1634  
QY 1545 TGAGACCCCTCTGTTTTCAGCTGACTGTGTTTGTACTTTCATGCTGGGATATTAATAGAAAAAG 1604  
Db 1635 TGAGACCCCTCTGTTCTGTGAACCTGTGAGCTCTACTTTCATGCTGGGTGTGAATTTCTAGGAC 1694  
QY 1605 TACAATATGTTGGTAGAATCGTGGGTGGAAACCAAGAAAAACAAGCTTACACCCCATATCAT 1664  
Db 1695 CAACACTCCTGTGGAGACGTGGAAAGGTTCCAAAGGCCAAACAGTCCCTATACCTACATCAT 1754  
QY 1665 CTTCAAGAAATGCAACTTTTACATTTACATGGGCATTCAGAGAACTAATCAGGGTCAAGA 1724  
Db 1755 TGAGGAGAACACTACCAGAGCTTCACCTGGGCCTTCCAGAGGACCACCTTTTCATGAGGC 1814  
QY 1725 TAATAGACGTTTCATCAATGACATGGTGAAGATTTATTCTATCACAGCCCACTAATGCAGT 1784  
Db 1815 AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCAACCAATGTTAT 1874  
QY 1785 TGATGGGTGGCGTCTCATGCCGTGCTGTGCCCTCGGTTCTGAACAGTGGGTTTCATC 1844  
Db 1875 GAATGGCGTGGCCTCCTACTGCCGTCCCTGTGCCCTAGAAGCCTCTGATGTGGGCTCCTC 1934  
QY 1845 GTGTGTCCCTGCCCTCCAGGCCACTACATTTGAGAAAGAAACCAACCAGTGCAAGGAATG 1904  
Db 1935 CTGCACCTCTGTCTCTGCTGTTACTATATTGACCGAGATTTCAGGAACCTGCCACCTCCTG 1994  
QY 1905 TCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTCCCATG 1964  
Db 1995 CCCCCCTAACACAATTCGAAAGCCCAACAGCTTATGGTGTCCAGGCCCTGTGTGCCCTG 2054  
QY 1965 CGGGCCTGGGAGTAAAAACAATCAGGACCAATTCGGTTTGGCTATAGTACTGCTTTTCTA 2024  
Db 2055 TGGTCCAGGGACCAAGAACAAAGATCCACTCTCTGTGTCTACAAATGATGACACCTTCTC 2114  
QY 2025 CCATGAAAAAGAAAATCAGATTTTGCACATATGACTTTTAGCAACCTCAGCAGTGTGGGCTC 2084  
Db 2115 ACGCAACACTCCAACACAGGACTTTCAACTACAACTTCTCCGCTTTGGCAAAACACCCGTAC 2174  
QY 2085 ATTAATGAATGGCCCCAGCTTCACCTCCAAGGAACAAAAATATCTCCATTTCTTCAATAT 2144  
Db 2175 TCTTGTGGAGGGCCAAAGCTTCACCTCCAAGGGTTGAAATACTTCCATCACTTTACCCCT 2234  
QY 2145 CAGTTTATGTGGCATGAGGGGAAGAGATGGCTCTCTGTACCAACAATATAACAGACTT 2204  
Db 2235 CAGTCTCTGTGGAACCAAGGTTAGGAAATGTCTGTGTGCACCCGACAATGTCACTGACCT 2294  
QY 2205 TACAGTAAAAGAAATAGTGGCAGGGTCAGATGATTACACAAAATTTGGTAGGGGCATTTGT 2264  
Db 2295 CCGGATTCTGAGGTGAGTCAGGG-----TTCTCCAAATCTATCACAGCCTACGT 2345  
QY 2265 ATGCCAGTCAACAATATTCTCTGAAAGTAAGGGTTCCGAGCAGCCTTATCATCACACA 2324  
Db 2346 CTGCCAGGCAGTCTATCATCCCCCAGAGGTGACAGGCTACAAAGGCCGGGTTTCTCTCACA 2405  
QY 2325 ATCCATCATTTCTGGCAGATACATTTCATAGGAGTTCACAGTTGAAACCACATTTGAAAAATAT 2384

Db 2406 GCCTGTACAGCCTTGTCTGATCGACTTATTGGGGTGACAACAGATATGACTCTGGATGGAAT 2465  
QY 2385 TAATATAAAAAGAGATATGTTTCCAGTTTCCAAACAAGCCAAAATACCAGATGTGCATTTCTT 2444  
Db 2466 CACCTCCCAGCTGAACCTTTTCCACCTGGAGTCTTGGGAATACCGGACGTGATCTTCTT 2525  
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAAATGGCCGATCAACTGCTGTGAAAAAT 2504  
Db 2526 TTATAGGTCCAATGATGTGACCCAGTCTCTGCAGTTCTGGGAGATCAACCACCTCCGCGT 2585  
QY 2505 GAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGTATTTCAGTCCCCAGCAAGTGCCCAGC 2564  
Db 2586 CAGGTGCAGTCCACAGAAAACTGTCCCTGGAAGTTTGTCTGCCAGGAACGTGCTCAGA 2645  
QY 2565 AGGTACCTGTGATGGGTGTACGTTCTATTTCCTGTGGGAGAGTGTGAAGCTTGGCCTCT 2624  
Db 2646 TGGGACCTGTGATGGCTGCAACTTCCACTTCCCTGTGGGAGAGCGGGCTGCTTGGCCGCT 2705  
QY 2625 GTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCTTCGAAGAGAGGATTTTCAGGAAC 2684  
Db 2706 CTGCTCAGTGGCTGACTACCATGCTATCGTCAGCAGCATGTGTGGTGGGATCCAGANGAC 2765  
QY 2685 CTTGTATGTGTGGAATGAACCTAAATGGTGCAATTAAGGAATTTCTTTGCCTGAGAAAAA 2744  
Db 2766 TACTTACGTGTGNCGAGAACCAAGCTATGCTCTGTGTGSCATTTCTCTGCCTGAGCAGAG 2825  
QY 2745 GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGGTGGGAGCCGGTGTGGGAGCTTT 2804  
Db 2826 AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG 2885  
QY 2805 TACTGCCGTTTGTGCTGGTCTGACCTGCTACTTCTGGAAGAAAGAAATCAAAAACTGGA 2864  
Db 2886 TACTGCCATCTGCTCACCGTCTGACCTGCTACTTTTGGAAAAAGAAATCAAAAACTAGA 2945  
QY 2865 ATACAAATATTCCAAGTTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCCGGCTGC 2924  
Db 2946 GTACAAGTACTCCAAGCTGGTGATGAATGCTACTCTCAAGGACTGTGACCTGCCAGCAGC 3005  
QY 2925 AGACAGTTGTGCTATCATGGAAGGAGAAGATAATGAAGAGGAAGTTGTATATTCCAAATAA 2984  
Db 3006 TGACAGCTGCGCCATCATGGAAGCGGAGGATGTAGAGGACGACCTCATCTTTACCAGCAA 3065  
QY 2985 ACAGTCACTACTAGGAAAACTCAAACTTTTGGCAACCAAGGAAAAAGAGACCATTTTGA 3044  
Db 3066 GAAGTCACCTTTTGGGAAGATCAAAATCATTTACCTCCAAGAGGACTCTGATGGATTGA 3125  
QY 3045 ATCTGTTCAACTGAAAAACCTC 3065  
Db 3126 CTCAGTGGCGCTGAAGACATC 3146

RESULT 10

US-10-121-049-37

; Sequence 37, Application US/10121049

; Publication No. US2003002239A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin





Db 2115' ACGCAACACTCCAAACCAGGACTTTCAACTACAACCTTCCCGCTTTGGCAAACACCCGTCAC 2174  
Qy 2085 ATTAATGAATGGCCCCAGCTTCACCTCCAAAGGAACAAATACTTCCATTTCTTCAATAT 2144  
Db 2175 TCTTGTGGAGGGCCAAAGCTTCACTTCCAAAGGGTTGAAATACTTCCATCACTTTACCT 2234  
Qy 2145 CAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTT 2204  
Db 2235 CAGTCTGTGGAAACCAGGGTAGGAAATGTCTGTGTGCACCGACAATGTCACTGACCT 2294  
Qy 2205 TACAGTAAAGAAATAGTGGCAGGTCAGATGATTACACAATTTGGTAGGGCATTTGT 2264  
Db 2295 CCGGATTCCTGAGGGTGAGTCAGG-----TTCTCCAAATCTATCACAGCCTACGT 2345  
Qy 2265 ATGCCAGTCAACAATATTCTCTGTGAAAGTAAGGGTTCCGAGCAGCCTTATCATCACA 2324  
Db 2346 CTGCCAGGAGTCATCATCCCCCAGAGGTGACAGGCTACAGGCCGGGTTTCCCTCACA 2405  
Qy 2325 ATCCATCATTTCTGGCAGATACATTTATAGGAGTCACAGTTTGAACACCATTTGAAAAATAT 2384  
Db 2406 GCCTGTCAGCCTTGCTGATCGACTTATTGGGGTGACAACAGATATGACTCTGGATGGAAT 2465  
Qy 2385 TAATATAAAGAAGATATGTTCCCGAGTTCACAACGCCAATACCAGATGTGCATTTCTT 2444  
Db 2466 CACCTCCAGCTGAACCTTTTCCACCTGGAGTCTCTTGGGAATACCGGACGTGATCTTCT 2525  
Qy 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAAATGGCCGATCAACTGCTGTGAAAT 2504  
Db 2526 TTATAGGTCCAATGATGTGACCCAGTCTGAGTCTGAGTCTGGGAGATCAACCCATCCGCGT 2585  
Qy 2505 GAGGTGTAATCTACTAAATCTGGAGCAGGAGTGATTTTCACTCCCGCAGCAAGTCCCCAGC 2564  
Db 2586 CAGGTGCAGTCCACAGAAAACCTGTCCCTGGAAAGTTTGCTGCTGCCAGGAACGTGCTCAGA 2645  
Qy 2565 AGGTACCTGTGATGGGTGTACGTTCTATTCTCTGTGGGAGAGTGCTGAAAGCTTGGCCCTCT 2624  
Db 2646 TGGACCTGTGATGGCTGCAACTTCCACTTCTCTGTGGGAGAGCGCGCTGCTTGGCCGCT 2705  
Qy 2625 GTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCTGCAAGAGAGGATTTCAAGAAAC 2684  
Db 2706 CTGCTCAGTGGCTGACTACCATGCTATCGTCAAGCAGCTGTGTGGCTGGGATCCAGANGAC 2765  
Qy 2685 CTTGTATGTGGAATGAACCTAAATGGTGCAATTAAGGAATTTCTTGGCTGAGAAAAA 2744  
Db 2766 TACTTACGTGTGCGAGAACCCCAAGCTATGCTGTGGTGGCATTTCTCTGCCTGAGCAGAG 2825  
Qy 2745 GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGGTGGGAGCGCGTGTGGGAGCTTT 2804  
Db 2826 AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG 2885  
Qy 2805 TACTGCCGTTTGTGGTGGCTCTGACCTGTGCTGACCTTCTGGAAGAAAGATCAAAAACTGGA 2864  
Db 2886 TACTGCCATCTGCTCACCGTCTTGACCTGTACTTTTGGAAAAAGAAATCAAAAACTAGA 2945  
Qy 2865 ATACAAATATCCAGTTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCCGGCTGC 2924  
Db 2946 GTACAAGTACTCCAAGCTGGTGATGAATGCTACTCTCAAGGACTGTGACCTGCCAGCAGC 3005  
Qy 2925 AGACAGTTGTGCTATCATGGAAGGAGAGATAATGAAGAGGAAGTTGTATATTCCAATAA 2984  
Db 3006 TGACAGCTGCGCCATCATGGAAGCGGAGGATGTAGAGGACGACCTCATCTTTACCAGCAA 3065  
Qy 2985 ACAGTCACTACTAGGAAACTCAAAATCTTTGGCAACCAAGGAAAAAGAGACCATTTTGA 3044  
Db 3066 GAAGTCACTTTTGGGAAGATCAAAATCATTTTACCTCCAAGAGGACTCTCTGATGGATTGA 3125  
Qy 3045 ATCTGTTCAACTGAAACCTC 3065  
Db 3126 CTCAGTGGCGCTGAAGACATC 3146

RESULT 11

US-10-123-904-37  
; Sequence 37, Application US/10123904  
; Publication No. US20030022328A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3330RIC54  
; CURRENT APPLICATION NUMBER: US/10/123,904  
; CURRENT FILING DATE: 2002-04-16  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 37  
; LENGTH: 3501  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 2762, 2778  
; OTHER INFORMATION: unknown base  
US-10-123-904-37

Query Match 27.8%; Score 987.8; DB 9; Length 3501;  
Best Local Similarity 60.5%; Pred. No. 1.7e-264;  
Matches 1755; Conservative 0; Mismatches 1119; Indels 27; Gaps 7;

Qy 171 GCTTCCTCCTTCCAGGAGAAAGATTATCACCTTGAATATACGGAATGTGATAGCAGTGG 230  
Db 267 GCTTCATGCCTGCAAGAGAGTCTGAGTACCACATATGAGTACAGCGCGTGTGACAGACGG 326  
Qy 231 CTCAGGTGGAGAGTTGCCATTCCAAATTTCTGAGTGGAGTCTTGGCCTGCCCTGACCC 290  
Db 327 TTCCAGGTGGAGGTGCGCGTGCGCATACCCCGGCTGTGCACAGCCTGTCTGACCC 386  
Qy 291 AGTGAGAGGCAAAAGAAATGCACCTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAA 350  
Db 387 CGTCAAGGCGCACCGAGTGTCTCTCTGCAACGCGCGGAGTTTCTGGATATGAAGGA 446  
Qy 351 CCAGGTATGCAGTAAGTGTGGTGAAGGCACCTATTCCCTTGGCAGTGGCATCAAAATTTGA 410  
Db 447 CCAGTCATGTAAGCCATGCGCTGAGGCGCGCTACTCCCTCGSCACAGGCATTCGGTTTGA 506  
Qy 411 TGAATGGGATGAATTCCCGGCAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGT 470  
Db 507 TGAGTGGGATGAGCTGCCCATGGCTTGGCAGCCTCTCAGCAACATGGAGCTGGATGA 566  
Qy 471 GGGCCCTTGTGACAGCAGGCGCAGCGGTGTAAACAACCTCTTCTTGGATCCCTCGTGGAAA 530  
Db 567 CAGTGTGTGAGTCCA---CCGGGAAGTGTACTTCTGTCCAAGTGGGTCCCCGGGGCGGA 623  
Qy 531 CTACATAGAATCTAATCGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAA 590  
Db 624 CTACATCGCCTCCAACACGCGACGAATGACAGCCACACTGATGATGACGCGCTCAACCTGAA 683  
Qy 591 GAAGTCAGGCTATGTCTTCTTTGAGTACCAGTATGTCGACAACAACATCTTCTTTGAGTT 650  
Db 684 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 743

QY	651	CTTTATTCAAAATGATCAGTCCAGGAGATGGACACCACTGACAAAGTGGGTAAACT	710
Db	744	TTTTCGTTCAGAAATGACCAGTCCAGCCCAATGCAGATGACTC--CAGGTGGATGAAGAC	800
QY	711	TACAGACAATGGAGAATGGGGCTCTCATTTCTGTAATGCTGAATCAGGCACAAACATACT	770
Db	801	CACAGAA--AGGATGGGAATCCACAGTGTGGAGCTAAATCGAGGCAATAATGTCTCT	857
QY	771	CTACTGGAGAACTACAGGCATCCTTATGGTTCTAAGGCGGTCAAGCCTGTGCTGGTAAA	830
Db	858	CTATTGGAGAACACAGCCTTCTCAGTATGGACCAAGTACCCAAAGCCTGTGCTGGTGA	917
QY	831	AAATATCACAATTTGAAGGGGTGGCGTACACATCAGAAATGTTTCTTGGCAAGCCAGC	890
Db	918	AAACATTGCCATAACAGGGGTGGCCTACACTTCAGAAATGCTTCCCCTGCAAACTGGC	977
QY	891	ATTACAGCAACAACACAGGTTTCATTCAACTGCCAGGTGTGTCCTCCAGAAACACCTATCTGA	950
Db	978	GTATGCAGACAAGCAGGGCTCTCTTTCTGCAAACTTTGCCCAGCCAACCTCTTATTCAA	1037
QY	951	GAAGGAGCCAAAGAATGTATAAGGTGTAA--AGACGACTCTCAATTTTCAGGATCCAG	1007
Db	1038	TAAAGGAGAAACTTCTTGCCACCAGTGTGACCCTGACAAATCTCAGAGAAAGGATCTTC	1097
QY	1008	TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAGAACTATTTCCAGATCCATACTCCATG	1067
Db	1098	TTCCTGTAACGTGCGCCACGCTGCACAGACAAGATTAATTTCTACACACACACGGCCTG	1157
QY	1068	TGATGAAGAAGGAAGACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGGA	1127
Db	1158	CGATGCCAACGGAGAGACAACTCATGTACAATGGGCCAAGCCGAAATCTGTAGCGA	1217
QY	1128	GGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAGAAGGATTGTCCGCCTTG	1187
Db	1218	GGACCTTGAGGGGCAGTGAAGCTGCCTGCCTCTGGTGTGAAGAGCCCACTGCCACCCCTG	1277
QY	1188	CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTCGAACAAT	1247
Db	1278	CAACCCAGGCTTCTTCAAAACCAACAACAGCACTGCCAGCCCTGCCCATATGGTTCCCTA	1337
QY	1248	TTACAGATGGAACCAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTT	1307
Db	1338	CTCCAATGG--CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGATT	1394
QY	1308	TGAATATAAATGGTGGAAATGTCTTCTTGCCATCCCTGTCTCTCTCGAATGTGG	1367
Db	1395	TGAATACAAATGGTGGAAACAGCTGCCCCACAACATGGAACGACCGTTCTCAGTGGGAT	1454
QY	1368	GAATTCAAAGTGGATGGAATGAATGGTTGGAGGTGGCTGGAGATCATATCCAGAGTGG	1427
Db	1455	CAACTTCGAGTACAAAGGCATGACAGGCTGGAGGTGGCTGGTGTACATTCACATTTACACAGC	1514
QY	1428	GGCTGGAGGTCTTGACAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAACC	1487
Db	1515	TGCTGGAGCCTCAGACAATGACTTTCATGATTCTCACTCTGGTGTGCCAGGATTTAGACC	1574
QY	1488	ACCAACATCTATGACTGGAGCCAC--GGTGTCTGAACTAGGAAGAATAAACATTTGTCTT	1544
Db	1575	TCCGCAGTCGGTGATGGCAGACACAGAGAATAAAGAGGTGGCCAGAAATCACAATTTGTCTT	1634
QY	1545	TGAGACCCCTCTGTTACGCTGACTGTGTTTGTACTTCATGGTGGATAATTAATAGAAAAAG	1604
Db	1635	TGAGACCCCTCTGTTCTGTGAACCTGTGAGCTCTACTTTCATGGTGGTGTGGAATTTCTAGGAC	1694
QY	1605	TACAAATGTGTAGAATCGTGGGTGGAAACCAAGAAACAAAGCTTACACCCATATCAT	1664
Db	1695	CAACACTCCTGTGGAGACGTGGAAAGGTTCCAAAGGCCAAACAGTCTCTATACCTACATCAT	1754
QY	1665	CTTCAAGAAATGCAACTTTTACATTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGA	1724
Db	1755	TGAGGAGAACAACACTACCACGAGCTTCACCTGGGCCCTTCCAGAGGACCACCTTTTCATGAGGC	1814

QY	1725	TAATAGAGCGTT	CATCAATGACATGCGTGAAGATT	TATCTATACAGACCAC	TAAATGCAGT	1784						
Db	1815	AAGCAGGAAGT	ACACCAATGACGTTGCCAAGAT	CTACTCCATCAATGT	CACCAATGTAT	1874						
QY	1785	TGATGGGTGGCGT	CCTCATGCCGTGCCCCCTCGGT	TCTGAACAGT	CGGGTTCATC	1844						
Db	1875	GAATGGCGTGGCC	TCTACTGCCGTCCCTGTGCCCT	TAGAAGCTCT	GATGTGGGCTCCTC	1934						
QY	1845	GTGTGTCCCCT	TGCCCTCCAGGCCACTACAT	TGAGAAAGAAACCA	ACCAGT	GCAAGGAATG	1904					
Db	1935	CTGCACCTCT	TGTCTCTGTGGT	TACTATATGAC	CCGAGAT	CAGGAACCTGCCACTCCCTG	1994					
QY	1905	TCCACCTGACACCT	TACCTGTCCATACATCAGGT	CTATGGCAAGAGAGGCT	TGTAT	TCCATG	1964					
Db	1995	CCCCCTTAACACAA	TCTGAAAGCCACACGCT	TATG	GTGTCCAGGCT	TGTGTGCCCTG	2054					
QY	1965	CGGGCTGGGAGT	AAAAACAATCAGGACCAAT	TGCGT	TGCTATAGT	GACTGCT	TGT	TCTA	2024			
Db	2055	TGGTCCAGGGACCA	GAACAACAAGATCCACT	CTCTGTGCTACAAT	GAT	TGCACCT	TCTC	2114				
QY	2025	CCATGAAAAAGAAA	TCAGAT	TGTCAC	TATGACT	TATAGCAACCT	TCAGCAGT	GTGGGCTC	2084			
Db	2115	ACGCAACACT	TCCAACCCAGGACT	TTCAACTACAAC	TCTCCGCT	TGCGCAACACACCGCT	CAC	2174				
QY	2085	ATTAATGAAT	GGCCCCCAGCT	TTCACCTCCAAAGGAACA	AAATACT	TCCAT	TCT	TCAATAT	2144			
Db	2175	TC	TGCTGGAGGGCCAA	GGCT	TCACTTCCAAAGGGT	TGAAAT	TACTTCCAT	CAC	T	2234		
QY	2145	CAG	T	TATGTGGGCAT	GAGGGGAAGAAGAT	GGCTCTCTGTACCAACA	ATATAACACAC	T	2204			
Db	2235	CAGT	CTCTGTGGAAACCAGGGT	AGGAAATGTCTGTGTGCACCCGACA	ATGTCACTGACCT	2294						
QY	2205	TACAGT	AAAAAGAAATAGTGGCAGGGT	TCAGATGAT	TACACAAAT	TGTTGGTAGGGGCAT	T	2264				
Db	2295	CCGGA	T	CTGAGGGTGAGT	CAAGG-----T	TCTCCAAAT	CTATCACAGCT	TACG	T	2345		
QY	2265	ATGCCAGT	CAACAATAT	TCCCTCTGAAAGTAAGGGT	TCCGAGCAGCCT	TATCAT	CACA	2324				
Db	2346	CTGCCAGGCAGT	CATCATCCCCCAGAGGT	GACAGGCTACAAGGCCGGGGT	TCTCTCACA	2405						
QY	2325	ATCCAT	CAT	TCTGGCAGAT	PACAT	TATAGGAGT	CAACAGT	TGAAACCACAT	T	2384		
Db	2406	GCCTGT	CAGCCT	TGCTGATCGACT	TATGGGGTGACAACAGAT	ATGACT	CTGGAT	GGAAAT	2465			
QY	2385	TAATATAAAGAA	GATATGTTCCAGT	TCCACCAAGCCAAAT	ACCAGAT	GTGCAT	T	2444				
Db	2466	CACCT	CCCCAGCTGA	ACTT	TCCACCTGGAGT	CC	TGGGAATACCGGAC	GTGAT	CTCT	2525		
QY	2445	TTATAAGT	CTTCTACAGCAACAACAT	CTTGTATTAATGGCCGAT	CAACTGCT	GTGAAAT	2504					
Db	2526	TTATAGGT	CCAATGATGTGACCCAGT	CTCGCAGT	CTGGGAGAT	CAACCAACAT	CCGCGT	2585				
QY	2505	GAGGT	GTAA	TCTACTAAATCTGGAGCAGGAT	GAT	T	CAGT	CCCCAGCAAGT	GCCCAGC	2564		
Db	2586	CAGGT	GCAGT	CCACAGAAAAC	TGTCCCTGGAA	TGTGCTGCCAGGA	CGT	GCT	CAGA	2645		
QY	2565	AGGT	ACCT	GTGATGGGT	TACGTTCTAT	TCCGTGGGAGAGT	GC	TGAAGCT	TGCCCT	2624		
Db	2646	TGGGACCT	GTGATGGCTGCAACT	TCCACT	TCTGTGGGAGAGCGG	GCT	T	GCCCCT	2705			
QY	2625	GTGT	ACGGAGCAT	GACT	TCCATGAGAT	TGAGGAGCCT	TGCAAGAGAGG	AT	TTCAGGAAAC	2684		
Db	2706	CTGCT	CAGTGGCT	GACT	TACCATGCT	ATCGT	CAGCAGCT	GTGTGGCT	GGGAT	CCAGANGAC	2765	
QY	2685	CTTGT	ATGTGTGGAAT	GAACCTAAATGGT	GCATTAAGGAAT	T	CTTGCCT	TGAGAAAAA	2744			
Db	2766	TACTT	ACGTGTGNCGAGA	ACCAAGCT	ATGCT	TGGTGGCAT	T	TCT	TGCCT	TGACGAG	2825	
QY	2745	GT	TGGCAACCT	GT	GAAACGGT	TGACT	T	TGGCTGAAGGT	GGGAGCCG	TGTGGGAGCT	T	2804
Db	2826	AGT	CACCAT	CTGCAAAACCA	TAGAT	T	TGGCTGAAAGT	GGGCAT	CTCT	TGCAGGCA	CTG	2885
QY	2805	TACT	GCCGTTTT	TGCTGGTGGCT	CTGACCT	GTCTGACCT	GTCTG	GGAAAAAGAA	T	CAAAAACT	TGGA	2864

```

Db 2886 TACTGCCATCCTGCTCACCGTCTTGACCTGCTACTTTTGGAAAAAGAAATCAAAAACTAGA 2945
QY 2865 ATACAAATATCCAAAGTTAGTAATGACGACTAACTCAAAGAGTGTGAACTCCCGGCTGC 2924
Db 2946 GTACAAGTACTCCAAGCTGGTGATGAATGCTACTCTCAAGGACTGTGACCTGCCAGCAGC 3005
QY 2925 AGACAGTTGTGCTATCATGGAAGGAGAGATAAATGAAGAGGAAGTTGTATATTCCAAATAA 2984
Db 3006 TGACAGCTGCGGCATCATGGAAGGCGAGGATGTAGAGGACGACCTCATCTTTACCAGCAA 3065
QY 2985 ACAGTCACTACTAGGAAAACTCAAACTCTTTGGCAACCAAGGAAAAAGAACCAATTTTGA 3044
Db 3066 GAAGTCACTTTTGGGAAGATCAAAATCATTTACCTCCAAGAGGACTCCTGTGATGGATTGA 3125
QY 3045 ATCTGTTCAACTGAAAAACCTC 3065
Db 3126 CTCAGTCCGCTGAAGACATC 3146
```

RESULT 12

```

US-10-140-470-37
; Sequence 37, Application US/10140470
; Publication No. US20030022331A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P33301C160
; CURRENT APPLICATION NUMBER: US/10/140,470
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 37
; LENGTH: 3501
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 2762, 2778
; OTHER INFORMATION: unknown base
US-10-140-470-37
```

```

Query Match 27.8%; Score 987.8; DB 9; Length 3501;
Best Local Similarity 60.5%; Pred. No. 1.7e-264;
Matches 1755; Conservative 0; Mismatches 1119; Indels 27; Gaps 7;

QY 171 GCTTCCTCCTTGCCAGGAGAAAGATTATACACTTTTGAATATACGGAATGTGATAGCAGTGG 230
Db 267 GCTTCATGCTGCAAGAGTCTGAGTACCACCTATGAGTACACGGCGGTGTGACAGCAGCGG 326
QY 231 CTCCAGGTGGAGAGTTGCCATTCCAAATCTGCAGTGGAGTGTCTTGGCCCTGCCGTGACCC 290
Db 327 TTCCAGGTGGAGGGTCGCCGTGCCGATACCCCGGCCCTGTGACACGACCTGTCTGACCC 386
QY 291 AGTGAGAGGCAAGAAATGCACATTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAA 350
```

```

Db 387 CGTCAAGGGCACCGAGTGTCTCTCTCTGCAACGCCGGGGAGTTTCTGGATATGAAGGA 446
QY 351 CCAGGTATGCAGTAAGTGTGGTGAAGGCACCTATTCCTTTGGCAGTGGCATCAAAATTTGA 410
Db 447 CCAGTATGTAAGCCATCGCTGAGGGCCGCTACTCCCTCGGCACAGGCATTCGGTTTGA 506
QY 411 TGAATGGGATGAATTGCGGCAGAGATTTTCTAACATTCGCAACATTCATGGACACTGTGGT 470
Db 507 TGAGTGGGATGAGTGCCTCCCATGGCTTTGCCAGCCTCTCAGCCAACATGGAGTGGATGA 566
QY 471 GGGCCTTCTGACAGCAGCGCAGAGCTGTAAACAACCTCTTCTTGGATCCCTCGTGGAAA 530
Db 567 CAGTGTCTGAGTCCA--CCGGGAAGTGTACTTCGTCCAAGTGGGTTCCTCCGGGGCGA 623
QY 531 CTACATAGAATCTAATCGTGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTAA 590
Db 624 CTACATCGCCTCCAACACGAGCAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 683
QY 591 GAAGTCAGGCTATGTCTTCTTGTAGTACCAGTATGTCGACAACAACATCTTCTTTGAGTT 650
Db 684 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 743
QY 651 CTTTATTCAAAATGATCAGTGCAGGAGATGGACACCCACCACTGACAAGTGGGTAAACT 710
Db 744 TTTTCGTTCAGAAATGACCAGTGCCAGCCCAATGCAGATGATC--CAGGTGGATGAAGAC 800
QY 711 TACAGACAATGGAGAAATGGGCTCTCTCATTTCTGTAATGCTGAATCAGGCACAAAACATCT 770
Db 801 CACAGAGAA--AGGATGGGAATTCACAGTGTGGAGCTAAATCGAGGCAATAATGTCTCT 857
QY 771 CTACTGGAGAACTACAGGCATCCTTATGGGTTCTAAGGCGGTCAAGCCTGTGCTGTAAA 830
Db 858 CTATTGGAGAACCAACAGCCTTCTCAGTATGGACCAGTACCCAAAGCCTGTGCTGGTGAG 917
QY 831 AAATATCACAATTGAAGGGTGGCGTACACATCAGAAATGTTTTCCTTGAAGCCAGGCAC 890
Db 918 AAACATTGCCATAACAGGGTGGCCTACACTTCAGAAATGCTTCCCTGCAAACTGGCAC 977
QY 891 ATTCAGCAACAACACCAGGTTCAATTCRACTGCCAGGTGTGTCCCAAGAAACACCTATTCTGA 950
Db 978 GTATGCAGACAAGCAGGGCTCCTCTTCTGTCAAACTTTGCCCAAGCAACTCTTATTCAA 1037
QY 951 GAAAGGAGCCAAAGAAATGTATAAGGTGTAA--AGACGACTCTCAATTTTCAGGATCCAG 1007
Db 1038 TAAAGGAGAAACTTCTTGCCACCAGTGTGACCCCTGACAAATACTCAGAGAAAGGATCTTC 1097
QY 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATCTCCATG 1067
Db 1098 TTTCTGTAACTGCGGCCAGCTTGCACAGACAAAAGATTATTTCTACACACACACGCGCTG 1157
QY 1068 TGATGAAGAAGGAAAGACACAGATAATGTACAAAGTGGATAGAGCCCAAAATCTGCCGGGA 1127
Db 1158 CGATGCCAACGGAGAGACACAACACTCATGTACAAATGGGCCAAGCCGAAATCTGTAGCGA 1217
QY 1128 GGATCTCACAGATGCTATTAGATTGCCCCCTTCTTGAGAGAGAAGAGATTGTCCGCCTTG 1187
Db 1218 GGACCTTGAGGGGSCAGTGAAGCTGCCCTGCCTCTGGTGTGAAGACCCCACTGCCACCCCTG 1277
QY 1188 CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTGGAACATTT 1247
Db 1278 CAACCCAGGCTTCTTCAAAACCAACAACAGCACCTGCCAGCCCTGCCCATATGTTCTCTA 1337
QY 1248 TTCAGATGGAACCAAGAATGTAGACCATGTCAGCAGGAAACGGAGCCCTGCACCTTGGCTT 1307
Db 1338 CTCCAATGG--CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAAACCTGTGTGGGATT 1394
QY 1308 TGAATATAAATGGTGAATGTCTCTTCTGGCAACATGAAAACCTTCTCTGCTTCAATGTTGG 1367
Db 1395 TGAATACAAATGGTGAACACAGCTGCCCAACAACATGGAACGACCCGTTCTCAGTGGGAT 1454
QY 1368 GAATTCAAAGTCCGATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427
Db 1455 CAACTTCGAGTACAAGGSCATGACAGGCTGGGAGGTGGCTGGTGTATCACATTTACACAGC 1514
```



QY 1428 GGCTGGAGGTTCTGACAATGATTACCTGATCTTTAAACTTGCATATCCCAGGATTTAAACC 1487  
Db 1515 TGCTGGAGCCTCAGACAATGACTTCATGATTTCTCACTCTGGTTGTGCCAGGATTTAGACC 1574  
QY 1488 ACCAACATCTATGACTGGAGCCAC---GGGTTCTGAACCTAGGAAGAATAACATTTTGTCTT 1544  
Db 1575 TCCGCAGTCCGTGATGGCAGACACAGAGAATAAAGAGGTGGCAGAAATCACATTTGTCTT 1634  
QY 1545 TGAGACCCCTCTGTTTCAGCTGACTGTGTTTGTACTTTCATGGTGGATATTAAATAGAAAAAG 1604  
Db 1635 TGAGACCCCTCTGTTCTGTGAACCTGTGAGCTCTACTTCATGGTGGGTGTAATTTCTAGGAC 1694  
QY 1605 TACAAATGGGTAGAAATCGTGGGTGGAAACCAAGAAAAACAAGCTTACACCCCATATCAT 1664  
Db 1695 CAACACTCCTGTGGAGACGTGGAAGGTTCCAAGGCCAAACAGTCC'TATACCTACATCAT 1754  
QY 1665 CTTCAAGAAATGCAAACTTTTACATTTTACATGGGCATTCAGAGAACTAAATCAGGGTCAAGA 1724  
Db 1755 TGAGGAGAACACTACCAGAGCTTCACCTGGGCCTTCCAGAGACCACCTTTTTCATGAGGC 1814  
QY 1725 TAATAGACGGTTTCATCAATGACATGGTGAAGATTTATTCTATCAGAGCCCACTAATGCAGT 1784  
Db 1815 AAGCAGGAAGTACACCAATGACGTTTGCCAAAGATCTACTCCATCAATGTCAACCAATGTTAT 1874  
QY 1785 TGATGGGGTGGCTCCTCANTGCCGTGCTGTCGCCCTCGGTTCTGAACAGTCCGGTTTCATC 1844  
Db 1875 GAATGGCGTGGCTCCTACTGCCGTCCCTGTGCCCTTAGAAGCCTCTGATGTGGGCTCCTC 1934  
QY 1845 GTGTGTCCCTGCCCTCCAGGCCACTACATTTGAGAAAGAAACCAACCAGTGCAGGAATG 1904  
Db 1935 CTGCACCTCTGTCTCTGTGTTACTATATTGACCGAGATTGACGGAACCTGCCACCTCCTG 1994  
QY 1905 TCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGAGGCTTGATTTCCATG 1964  
Db 1995 CCCCCCTAACACAATTTCTGAAAGCCCAACAGCCTTATGGTGTCCAGGCCGTGTGGCCCTG 2054  
QY 1965 CGGGCCTGGAGTAAAAACAATCAGSACCATTTCGGTTTGGCTATAGTGACTGCTTTTCTTA 2024  
Db 2055 TGGTCCAGGACCAAGAACAACAAGATCCCACTCTCTGTGCTACAAATGATTGCACCTTCTC 2114  
QY 2025 CCATGAAAAAGAAATCAGATTTTTCGACTATGACTTTTAGCRAACCTCAGCAGTGTGGGCTC 2084  
Db 2115 ACGCAACACTCCAAACCAGGACTTTCAACTACAACCTTCTCCGGTTTGGCAACACCCGTAC 2174  
QY 2085 ATTAATGAATGGCCCCCAGCTTCACCTCCAAAGGAACAAATACTTCCATTTCTTCAATAT 2144  
Db 2175 TCTTGTGGAGGGCCCAAGCTTCACTTCCAAAGGGTTGAAATACTTCCATCACTTTACCCCT 2234  
QY 2145 CAGTTTATGTGGGCATGAGGGGAAGAGATGGCTCTCTGTACCAACAATATAACAGACTT 2204  
Db 2235 CAGTCTCTGTGGAACCCAGGGTAGGAAATGTCTGTGTGCACCGACAATGTCACTGACCT 2294  
QY 2205 TACAGTAAAGAAATAGTGGCAGGGTCAAGATGATTACACAAATTTGGTAGGGGCATTTGT 2264  
Db 2295 CCGGATTCCGTAGGGTGAGTCAGGG-----TTCTCCAAATCTATCACAGCCTACGT 2345  
QY 2265 ATGCCAGTCAACAATTATTCCTTCTGAAAAGTAAGGGTTTCCGAGCAGCCTTATCATCACA 2324  
Db 2346 CTGCCAGGCATCATATCCCCCAGAGGTGACAGGCTACAAAGGCCGGGTTTCCCTCACA 2405  
QY 2325 ATCCATCATTTCTGGCAGATACATTTCATAGGAGTCACAGTTTGAACACCACATTTGAAAAATAT 2384  
Db 2406 GCCTGTCAGCTTGCTGATCGACTTATTGGGGTGACAACAGATATGACTCTGGATGGAT 2465  
QY 2385 TAATATAAAGAAGATATGTTCCCACTTCCAACAAGCCCAATACCAGATGTGCATTTCTT 2444  
Db 2466 CACCTCCCAGCTGAACCTTTTCCACCTGGAGTCCCTTGGGAATACCCGACGTGATCTTCTT 2525  
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAAATGGCCCATCAACTGCTGTGAAT 2504  
Db 2526 TTATAGGTCCAATGATGTGACCCAGTCTCTGCAGTTCTGGGAGATCAACCAACCATCCGGCT 2585

QY 2505 GAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCCCAGCAAGTGCCCCAGC 2564  
Db 2586 CAGGTGCAGTCCACAGAAAACTGTCCCTGGAAGTTTGCTGCTGCCAGGAACGTGCTCAGA 2645  
QY 2565 AGGTACCTGTGA'GGGTGTACGTTCTATTTCCCTGTGGGAGAGTGTGTAAGCTTGCCCTCT 2624  
Db 2646 TGGGACCTGTGATGGCTGCAACTTCCACTTCC'TGTGGGAGAGCGCGGCTGCTGCCGCT 2705  
QY 2625 GTGTACGGGAGCATGACTTCCATGAGAT'GTAGGGAGCCTTCAAAGAGAGGATTTTCAGGAAC 2684  
Db 2706 CTGCTCAGTGGCTGACTACCATGCTATCGTCAAGAGCTGTGTGGCTGGGATCCAGANGAC 2765  
QY 2685 CTTGTATGTGTGGAATGAACCTAAATGGTGCAT'TAAGGAATTTCTTTGCCTGAGAAAAA 2744  
Db 2766 TACTTACGTGTGNCGAGAACCCAAAGCTATGCTCTGGTGGCATTTCTCTGCCTGAGCAGAG 2825  
QY 2745 GTTGGCAACCTGTGAAACGGT'GACTTTTGGCTGAAGGTGGGAGCCGGTGTGGGAGCTTT 2804  
Db 2826 AGTCACCATCTGCAAAACCATAGATTTTCTGGCTGAAGGTGGGCATCTCTGCAGGCACCTG 2885  
QY 2805 TACTGCCGTTT'TGGTGGCTCTGACCTGCTACTTCTGGAAAGAAATCAAAAACTTGA 2864  
Db 2886 TACTGCCATCCTGCTCACCGTCTTGACCT'GCTACTTTTGGAAAAAGAAATCAAAAACTAGA 2945  
QY 2865 ATACAAATAT'TCCAAGTTAGTAATGACGACTAACTCAAAAGAGTGTGAAC'TCCCGGCTGC 2924  
Db 2946 GTACAAGTACT'CCAAGCTGGT'GATGAATGCTACT'CTCAAGGACT'GTGACCT'GCCAGCAGC 3005  
QY 2925 AGACAGTTGTGCTATCATGGAAGGAGAAGATAAT'GAAGAGGAAGTTTGTATAT'TCCANTAA 2984  
Db 3006 TGACAGCTGCGCCATCATGGAAGGCGAGGATGTAGAGGACGACCTCATCTTTACCAGCAA 3065  
QY 2985 ACAGTCACTACTAGGAAAACTCAAATCTTTGGCAACCAAGGAAAAAGAACCACTTTTGA 3044  
Db 3066 GAAGTCACTTTTGGGAAGATCAAATCAITTTACCTCCAAGAGGACTCCTGATGGATTTGA 3125  
QY 3045 ATCTGTTCAACTGAAAACTC 3065  
Db 3126 CTCAGTGGCGCTGAAGACATC 3146

RESULT 13

US-10-175-746-37

; Sequence 37, Application US/10175746

; Publication No. US20030027270A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330RIC353

; CURRENT APPLICATION NUMBER: US/10/175,746

; CURRENT FILING DATE: 2002-06-19

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 37

; LENGTH: 3501

; TYPE: DNA

; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 2762, 2778
; OTHER INFORMATION: unknown base
US-10-175-746-37

Query Match 27.8%; Score 987.8; DB 9; Length 3501;
Best Local Similarity 60.5%; Pred. No. 1.7e-264;
Matches 1755; Conservative 0; Mismatches 1119; Indels 27; Gaps 7;

Qy	171	GCTTCCTCCTGCCAGGAGAAAGATTATCACTTTTGAATATACGGAATGTGATAGCAGTGG	230
Db	267	GCTTCATGCCTGCAAGAGTCTGAGTACCACCTATGAGTACACGGCGTGTGACAGCACGGG	326
Qy	231	CTCCAGGTGGAGAGTTGCCATTCCAAATCTGCAGTGGACTGCTTGGCCTGCCTGACCC	290
Db	327	TTCCAGGTGGAGGGTCGCCGTGCCGATACCCCGGCCCTGTGCACCAGCCTGTCTGACCC	386
Qy	291	AGTGAGAGGCAAGAATGCACCTTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAA	350
Db	387	CGTCAAGGGCACCGAGTGTCTCTCTGCAACGCCGGGGAGTTTCTGGATATGAAGGA	446
Qy	351	CCAGGTATGCAGTAAAGTGTGGTGAAGGCACCTTATTCCTTTGGGCAGTGGCATCAAAATTGA	410
Db	447	CCAGTCATGTAAGCCATGCGCTGAGGGCGCGTACTCCCTCGGCACAGGCATTCGGTTTGA	506
Qy	411	TGAATGGGATGAATGCCGGCAGGATTTTCTAAACATCGCAACATTCATGGACACTGTGGT	470
Db	507	TGAGTGGGATGAGTCCCCCATGGCTTTGGCAGCCCTCTCAGCCAAACATGGAGCTGGATGA	566
Qy	471	GGGCCCTTCTGACAGCAGGCCAGACGGCTGTAAACAACTCTTCTTGGATCCCTCGTGGAAA	530
Db	567	CAGTGTCTGTAGTCCA---CCGGGAACGTGTACTTCGTCCAAGTGGGTTCCTCCGGGGCGA	623
Qy	531	CTACATAGAATCTAATCGTGATGACTGCACGGTGTCTTTGATCTATGCTGTCACCTTAA	590
Db	624	CTACATCGCCTCCACACAGGACGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA	683
Qy	591	GAAGTCAAGGCTATGTCTTTGAGTACCAGTATGTGCGACAACAACATCTTCTTTGAGTT	650
Db	684	GCAATCTGGCACCCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT	743
Qy	651	CTTTATTCAAAATGATCAGTCCAGGAGATGGACACCACCACCTGACAAAGTGGGTAACACT	710
Db	744	TTTCGTTCAGAAATGACCAGTGCCAGCCCAATGCAGATGACTC---CAGGTGGATGAAGAC	800
Qy	711	TACAGACAATGGAGAAATGGGGCTCTCATTTCTGTAAATGCTGAAATCAGGCACAAACATACT	770
Db	801	CACAGAGAA---AGGATGGGAATTCACAGTGTGGAGCTAAATCGAGGCAATAATGTCTCT	857
Qy	771	CTACTGGAGAACTACAGGCATCCTTATGGGTTCTAAGCGGGTCAAGCCTGTGCTGGTAAA	830
Db	858	CTATTGGAGAACCAACAGCCTCTCAGTATGGACCAAAAGTACCCAAGCCTGTGCTGGTGAG	917
Qy	831	AAATATCACAATTGAAGGGTGGGTACACATCAGAATGTTTTCTTGCAGGCCAGGCAC	890
Db	918	AAACATTGCCATAACAGSGGTGGCCTACACTTCAGAATGCTTCCCTTGCAAACTTGGCAC	977
Qy	891	ATTACGACAACAACCAGGTTCAATTCAACTGCCAGGTGTGTCCCAGAAACACCTATTCTGA	950
Db	978	GTATGCAGACAAGCAGGGCTCTCTTTCTGCAAACTTTGCCAGGCCAACTCTATTCAA	1037
Qy	951	GAAGGAGCCAAAGAATGTATAAGGTGTA---AGACGACTCTCAATTTTTCAGGATCCAG	1007
Db	1038	TAAAGGAGAAACTTCTTGCCACCAGTGTGACCCCTGACAAATACTCAGAGAAAGGATCTTC	1097
Qy	1008	TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATACTCCATG	1067
Db	1098	TTCTCTGTAACTGCGCCAGCTTGACAGACAAGAGTATTTCTACACACACACGCGCTG	1157
Qy	1068	TGATGAAGAAGGAAGACACAGATAAATGTACAAAGTGGATAGAGCCCCAAAATCTGCGGGGA	1127

Db	1158	CGATGCCAACGGAGAGACACAACACTCATGTACAAATGGGCCAAGCCGAAAATCTGTAGCGA	1217
Qy	1128	GGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAGAAGAGATTGTCCGCCCTTG	1187
Db	1218	GGACCTTGAGGGGGCAGTGAAGCTGCCTGCCTCTGGTGTGAAGACCCCACTGCCCAACCCTG	1277
Qy	1188	CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCCTCGGAACATT	1247
Db	1278	CAACCCAGGCTTCTTCAAAACCAACAACAGCACCTGCCAGCCCTGCCCATATGGTTCTCTA	1337
Qy	1248	TTCAGATGGAACCAACAAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTGGCTT	1307
Db	1338	CTCCAATGG--CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGATT	1394
Qy	1308	TGAATATAAATGGTGGAAATGTCTTCTTGGCAACATGAAAAACTTCTCTGCTTCAATGTTGG	1367
Db	1395	TGAATACAAATGGTGGAAACAGCTGCCCAACAACATGAAACGACCGTTCTCAGTGGGAT	1454
Qy	1368	GAAATCAAAGTCCGATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGG	1427
Db	1455	CAACTTCGAGTACAAGGGCATGACAGGCTGGGAGGTGGTGGTATCACATTTACACAGC	1514
Qy	1428	GGCTGGAGGTTCTGACAAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTAAACC	1487
Db	1515	TGCTGGAGCCTCAGACAAATGACTTCATGATTCTCACTCTGTTGTGCCAGGATTAGACC	1574
Qy	1488	ACCAACATCTATGACTGGAGCCAC--GGGTTCTGAACTAGGAAGAATAACATTTGTCTT	1544
Db	1575	TCCGAGTCGGTGATGGCAGACACAGAGAATAAAGAGTGGCCAGAATCACATTTGTCTT	1634
Qy	1545	TGAGACCCCTCTGTTTCAGCTGACTGTGTTTTGTACTTTCATGGTGGATATTAATAGAAAAAG	1604
Db	1635	TGAGACCCCTCTGTTCTGTGAACGTGAGCTCTACTTTCATGTTGGGTGTGAATTTCTAGGAC	1694
Qy	1605	TACAAATGTGGTAGAATCGTGGSGTGGAAACCAAGAAAACAAGCTTACACCCATATCAT	1664
Db	1695	CAACACTCTCTGTGGAGACGTGGAAGGTTCCAAAGGCAACAGTCCCTATACCTACATCAT	1754
Qy	1665	CTTCAAGAATGCAACTTTTACATTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGA	1724
Db	1755	TGAGGAGAACACTACCACGAGCTTCACCTGGGCCCTCCAGAGGACCACCTTTTCATGAGGC	1814
Qy	1725	TAATAGACGGTTCATCAATGACATGGTGAAGATTTATTCTATCACAGCCCACTAATGCAGT	1784
Db	1815	AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCAACCAATGTTAT	1874
Qy	1785	TGATGGGTGGCGTCCCTCATGCGCTGCCTGTGCCCTCGGTTCTGAACAGTCGGGTTCTATC	1844
Db	1875	GAATGGCGTGGCCTCCTACTGCCGTCCCTGTGCCCTAGAAAGCCTCTGATGTGGGCTCCTC	1934
Qy	1845	GTGTGTCCCTGCCCTCCAGGCCACTACATTTGAGAAAAGAAACCAACCACTGCAAGGAATG	1904
Db	1935	CTGCACCTCTGTCTCTGTTACTATATTGACCGAGATTTCAGGAACCTGCCACTCCTG	1994
Qy	1905	TCCACCTGCACACTACCTGTCCATACATCAGGTCTATGGCAAAGAGGCTTGTATTCCATG	1964
Db	1995	CCCCCTAACACAATCTGAAAGCCCCACAGCCTTATGGTGTCCAGGCTGTGTGCCCTG	2054
Qy	1965	CGGGCCTGGGAGTAAAAACAATCAGGACCATTTCGGTTTGTCTATAGTACTGCTTTTCTA	2024
Db	2055	TGGTCCAGGGACCAAGAACAAGATCCACTCTCTGTGCTACAAATGATTGCACCTTCTC	2114
Qy	2025	CCATGAAAAAGAAATCAGATTTTGCATATGACTTTTAGCAACCTTCAGCAGTGTGGGCTC	2084
Db	2115	ACGCAACACTCCAACCCAGGACTTTTCAACTACAACTTCTCCGCTTTGGCAAAACACCGTCA	2174
Qy	2085	ATTAATGAATGGCCCCAGCTTCACCTCCAAAGGAACAAAATACTTCCATTTCTTCAATAT	2144
Db	2175	TCTTGTGAGGGCAAGCTTCACCTTCCAAAGGGTTGAAATACTTTCATCACTTTTACCCT	2234
Qy	2145	CAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTT	2204
Db	2235	CAGTCTCTGTGGAAACCAGGGTAGGAAAATGTCTGTGCAACCGACAATGTCACTGACCT	2294

QY 2205 TACAGTAAAGAAATAGTGGCAGGGTCAGATGATTACACAAATTTGGTAGGGGCATTGT 2264  
Db 2295 CCGGATTCCTGAGGGTCAGTCAGG-----TTCTCCAAATCTATACAGCCTACGT 2345  
QY 2265 ATGCCAGTCAACAATATATCTCTGAAAGTAAGGGTTTCCGAGCAGCCTTATCATCACA 2324  
Db 2346 CTGCCAGGCAGTCATCATCCCCCAGAGGTCAGAGGCTACAAGGCCCGGGTTTCTCTACA 2405  
QY 2325 ATCCATCATTTCTGGCAGATACATTTCCAGTTCCACAGTTCAGAAACCACATTTGAAAAATAT 2384  
Db 2406 GCCTGTCAGCCTTGGTGCAGCTTATTTGGGTGACAACAGATATGACTCTGGATGGAAT 2465  
QY 2385 TAATATAAAGAGATATGTTCCAGTTTCCACAGTTCCACAAAGCCAAATACCAGATGTGCATTTCTT 2444  
Db 2466 CACCTCCCAGCTGAACCTTTCCACCTGGAGTCTTGGGAATACCGACGTGATCTTCTT 2525  
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAATGGCCGATCAACTGCTGTGAAAAAT 2504  
Db 2526 TTATAGGTCCAATGATGTGACCCAGTCTCGCAGTTCTGGGAGATCAACCACCATCCGCGT 2585  
QY 2505 GAGGTGTAATCTACTAAATCTGGAGCAGGAGTGAATTCAGTCCCCAGCAAGTGCACG 2564  
Db 2586 CAGGTGCAGTCCACAGAAAACTGTCCCTGGAAGTTTGCTGCTGCCAGGAACGTGCTCAGA 2645  
QY 2565 AGGTACCTGTGATGGGTGTACGTTCTATTTCCCTGTGGGAGAGTGTGAAAGCTTTGCCCTCT 2624  
Db 2646 TGGGACCTGTGATGGCTGCAACTTCCACTTCCCTGTGGGAGAGCGCGCTGTGCCCGCT 2705  
QY 2625 GTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCTGCAAGAGAGAGGATTTCAGGAAAC 2684  
Db 2706 CTGCTCAGTGGCTGACTACCATGCTATCGTCAAGCAGCTGTGTGGTGGATCCAGANGAC 2765  
QY 2685 CTTGTATGTGGAATGAACCTAAATGGTGCATTAAAGGAATTTCTTTGCCTGAGAAAAA 2744  
Db 2766 TACTTACGTGTGCGAGAACCCAGCTATGCTCTGGTGGCATTTCTCTGCCTGAGCAGAG 2825  
QY 2745 GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGGTGGGAGCGCGGTGTGGGAGCTTT 2804  
Db 2826 AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG 2885  
QY 2805 TACTGCCGTTTGTGCTGGTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATCAAAAACCTGA 2864  
Db 2886 TACTGCCATCTGCTCACCGTCTTGACCTGCTACTTTTGGAAAAAGAAATCAAAAACCTAGA 2945  
QY 2865 ATACAAATATTTCCAAGTTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCCGGCTGC 2924  
Db 2946 GTACAAGTACTTCCAAGCTGGTGTGATGAATGCTACTCTCAAGGACTGTGACCTGCCAGCAG 3005  
QY 2925 AGACAGTTGTGCTATCATGGAAGGAGAGAGATANTGAAGAGGAAGTTGTATATTCCAATAA 2984  
Db 3006 TGACAGCTGGCCCATCATGGAAGGCGAGGATGTAGAGGACGACCTCATCTTTACCAGCAA 3065  
QY 2985 ACAGTCACTACTAGGAAAACTCAAAATCTTTGGCAACCAAGGAAAAAGAACCATTTTGA 3044  
Db 3066 GAAGTCACCTTTTGGGNAGATCAAAATCATTTACCTCCAAGAGGACTCCTGTGATGATTGA 3125  
QY 3045 ATCTGTTCAACTGAAAAACCTC 3065  
Db 3126 CTCAGTGCCGCTGAAGACATC 3146

RESULT 14

US-10-176-918-37

; Sequence 37, Application US/10176918  
; Publication No. US20030027275A1

GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3330R1C382  
; CURRENT APPLICATION NUMBER: US/10/176,918  
; CURRENT FILING DATE: 2002-06-20  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 37  
; LENGTH: 3501  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: 2762, 2778  
; OTHER INFORMATION: unknown base  
US-10-176-918-37

Query Match 27.8%; Score 987.8; DB 9; Length 3501;  
Best Local Similarity 60.5%; Pred. No. 1.7e-264;  
Matches 1755; Conservative 0; Mismatches 1119; Indels 27; Gaps 7;

QY 171 GCCTTCTCTTGGCCAGGAGAAAGATTATCACATTTGAATATACGGAATGTGATACGAGTGG 230  
Db 267 GCCTCATGCTGCAAAAGAGTCTGAGTACCACATATGAGTACACGGCGTGTGACAGCACGGG 326  
QY 231 CTCAGGTGGAGAGTTGCCATTCCAAATTTCTCCTGTGCTTCTGAGTGGACTGCTCTGGCCCTGCCTGACCC 290  
Db 327 TTCCAGGTGGAGGGTCGCCGTGCCGCATACCCCGGGCCTGTGCACCAGCCTGTCTGACCC 386  
QY 291 AGTGAGAGGCAAGAAATGCACATTTCTCCTGTGCTTCTGAGAGAGTATCTAGAAATGAAGAA 350  
Db 387 CGTCAAGGGCACCCGAGTGTCTTCTCTCTGCAACGCCGGGAGTTTCTGGATATGAAGGA 446  
QY 351 CCAGGTATGCAGTAAGTGTGGTGAAGGCACCTATTTCTTGGGAGTGGCAGTGCATCAAAATTTGA 410  
Db 447 CCAGTCATGTAAGCCATGCCGTGAGGGCCGCTACTCCCTCGGCACACAGGCATTCGGTATTGA 506  
QY 411 TGAATGGGATGAATTTGCCGCGCAGGATTTTCTTAACATCGCAACATTCATGGACACTGTGGT 470  
Db 507 TGAGTGGGATGAGCTGCCCATGGCTTTGCCAGCCTCTCAGCCAACATGGAGCTGGATGA 566  
QY 471 GGGCCCTTCTGACAGCAGCCAGACGGCTGTAAACAACTCTTCTTGGATCCCTCGTGGAAA 530  
Db 567 CAGTGTCTGTAGTCCA--CCGGGAACGTACTTCTGTCCAAGTGGGTTCGCCCGGGGCGA 623  
QY 531 CTACATAGAATCTAATCGTGCATGACTGCACGGTGTCTTGTATCTATGCTGTGCACCTTAA 590  
Db 624 CTACATCGCCTCCAACACGACGACGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 683  
QY 591 GAAGTCAGGCTATGTCTTCTTGTAGTACCAGTATGTCGACAAACAACATCTTCTTTGAGTT 650  
Db 684 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 743  
QY 651 CTTTATTTCAAAATGATCAGTCCAGGAGATGGACACCACCACTGACAAAGTGGGTAAAAC 710  
Db 744 TTTCGTTCAGAAATGACCAAGTCCAGGCCCAATGCAGATGACTC---CAGGTGGATGAAGAC 800  
QY 711 TACAGACAATGGAGAAATGGGGCTCTCATTTCTGTAAATGCTGAAATCAGGCAACAAACATCT 770  
Db 801 CACAGAGAA--AGGATGGGAATTCACACAGTGTGGAGCTAAATCGAGGCAATAATGTCTCT 857  
QY 771 CTACTGGAGAACTACAGGCATCTTATGGGTTCTAAGGGGCTCAAGCCCTGTGCTGGTAAA 830



Db 858 CTATTGGAGAACACAGCCTTCTCAGTATGGACCAAAAGTACCCAAAGCCTGTGCTGGTGAG 917  
QY 831 AAATATCACAAATTGAAGGGGTGGCGTACACATCAGAATGTTTTCCTTGCAGCCAGGCAC 890  
Db 918 AAACATTTGCCATAACAGGGGTGGCCTACACTTCAGAATGCTTCCCCTGCAAAAGCTGGCAC 977  
QY 891 ATTACAGCAACAAACAGGTTTCATTCAACTGCCAGGTGTGTCCTCCAGAAACACCTTATCTGA 950  
Db 978 GTATGCAGACAAGCAGGGCTCCTTCTTGCAAACTTTGCCAGCCAACTCTTATTCAAA 1037  
QY 951 GAAAGGAGCCAAAGAAATGTATAAGGTGTAA--AGACGACTCTCAATTTTCAGGATCCAG 1007  
Db 1038 TAAAGGAGAACTTCTTGCCACCAGTGTGACCCCTGACAAATACTCAGAGAAAGGATCTTC 1097  
QY 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAGACTATTTCCAGATCCCATG 1067  
Db 1098 TTCTGTAAAGTGGCCCGAGCTTGCACAGACAAGATTATTCTACACACACACGGCCTG 1157  
QY 1068 TGATGAAGAGGAAAGACACAGATAATGTACAAAGTGGATAGAGCCCAAAATCTGCCGGGA 1127  
Db 1158 CGATGCCAACGGAGAGACACAACACTCATGTACAAATGGGCCAAGCCGAAATCTGTAGCGA 1217  
QY 1128 GGATCTCACAGATGCTATTAGATTGCCCCCTTCTTGCCATCCCTGTCTCCTGGAACAT 1187  
Db 1218 GGACCTTGAGGGGGCAGTGAAGCTGCCTGCTCTGGTGTGAAGACCCACTGCCCAACCTG 1277  
QY 1188 CAACCCCTGGAATTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCCTGGAACAT 1247  
Db 1278 CAACCCAGGCTTCTTCAAAACCAACAACAGCACCTGCCAGCCCTGCCCATATGGTTCCTA 1337  
QY 1248 TTCAGATGGAACCAAGAAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTT 1307  
Db 1338 CTCCAATGG--CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGATT 1394  
QY 1308 TGAATATAAATGGTGGAAATGTCCTTCCCTGGCAACATGAAAACCTTCCCTGCTTCAATGTTGG 1367  
Db 1395 TGAATACAAATGGTGGAAACACGCTGCCCAACAACATGGAACGACCGTCTCTCAGTGGGAT 1454  
QY 1368 GAATTCAAAGTCGATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427  
Db 1455 CAACCTCGAGTACAAAGGCATGACAGGCTGGGAGGTGGCTGGTGATCACATTTACACAGC 1514  
QY 1428 GGCTGGAGGTTCTGACAATGATTACCTGATCTTAACTTGCATATCCAGGATTTAAACC 1487  
Db 1515 TGCTGGAGCCTCAGACAATGACTTTCATGATTTCTCACTCTGGTTGTGCCAGGATTTAGACC 1574  
QY 1488 ACCAACATCTATGACTGGAGCCAC--GGGTTCTGAACTAGGAAGATTAACATTTGTCTT 1544  
Db 1575 TCCGCAGTCGGTATGGCAGACACAGAGAATAAAGAGGTGGCCAGAAATCACATTTGTCTT 1634  
QY 1545 TGAGACCCCTCTGTTACGCTGACTGTGTTTGTACTTCATGGTGGATATTAAAGAAAAG 1604  
Db 1635 TGAGACCCCTCTGTTCTGTGAACCTGTGAGCTCTACTTCATGGTGGTGTGAATTCATAGGAC 1694  
QY 1605 TACAAATGGTAGAATCGTGGGTGGAAACCAAGAAAACAAAGCTTACACCATATCAT 1664  
Db 1695 CAACACTCCTGTGGAGACGTGGAAGGTTCCAAAGGCAAAACAGTCCATATACCTACATCAT 1754  
QY 1665 CTTCAAGAATGCAACTTTTACATTTACATGGGCATTCAGAGAACTAATCAGGGTCAAGA 1724  
Db 1755 TGAGGAGAACACTACCACGAGCTTCACCTGGGCCCTTCAGAGGACCACCTTTTCATGAGGC 1814  
QY 1725 TAATAGACGGTTTCATCAATGACATGGTGAAGATTTTATCTATCACAGCCCACTAATGCAGT 1784  
Db 1815 AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCACCAATGTTAT 1874  
QY 1785 TGATGGGTGGCGTCCCTCATGCGGTGCCCTGTGCCCTCGGTTCTGAACAGTCGGGTTCATC 1844  
Db 1875 GAATGGCGTGGCCCTCCTACTGCGCTCCCTGTGCCCTAGAACCTCTGTATGTGGGCTCCTC 1934  
QY 1845 GTGTGTCCCTGCCCTCCAGGCCCACTACATTTGAGAAAGAAACCAACCAAGTGCAGGAATG 1904

Db 1935 CTGCACCTCTTGTCTCTGCTGGTTACTATATTGACCGAGATTTCAGGAACCTGCCACTCCTG 1994  
QY 1905 TCCACCTGACACCTACCTGTCCATACATACATCAGGTCTATGGCAAGAGAGGCTTGTATTCATG 1964  
Db 1995 CCCCCCTAACACAATCTGAAAGCCCAACAGCCTTATGGTGTCCAGGCCTGTGTGCCCTG 2054  
QY 1965 CGGGCCTGGAGTAAAAAACAAATCAGGACCAATTCGGTTTGTCTATAGTACTGCTTTTCTA 2024  
Db 2055 TGGTCCAGGGACAAAGAACAAAGATCCACTCTCTGTGTACATGATTCACCTTCTC 2114  
QY 2025 CCATGAAAAAGAAATCAGATTTTGCACATGACTTTTAGCAACCTCAGCAGTGTGGGCTC 2084  
Db 2115 ACGCAACACTCCAACACGAGACTTTCAACTACAACCTTCTCCGCTTTGGCAAAACACCGTCAC 2174  
QY 2085 ATTAATGAATGGCCCCAGCTTCACCTCCAAGGAACAAATACTTCCATTTCTCAATAT 2144  
Db 2175 TCTTGTGGAGGGCCAGCTTCACITCCAAAGGTTGAAATACTTCCATCACTTTACCT 2234  
QY 2145 CAGTTTATGTGGGCATGAGGGGAAGATGGCTCTCTGTACCAACAATATAACAGACTT 2204  
Db 2235 CAGTCTCTGTGGAACACGAGGTAGGAAATGTCTGTGCACCGACAATGTCACCTGACCT 2294  
QY 2205 TACAGTAAAGAAATAGTGGCAGGGTCAGATGATTACACAAATTTGGTAGGGGCATTGT 2264  
Db 2295 CCGGATTCTGAGGGTGAGTCAGGG-----TCTCCAAATCTATCACAGCCTACGT 2345  
QY 2265 ATGCCAGTCAACAAATTAATCTCTGAAAGTAAGGTTTCCGAGCAGCCTTATCATCACA 2324  
Db 2346 CTGCCAGGCAGTCATCATCCCCCAGAGGTGACAGGTACAGGCTACAAGCCGGGTTTCCTCACA 2405  
QY 2325 ATCCATCATTTCTGGCAGATACATTCATAGGAGTCAAGTTGAAACCAACATTTGAAAAATAT 2384  
Db 2406 GCCTGTCAGCCTTGTGTGATCGACTTATTTGGGTGACAACAGATATGACTCTGGATGGAAT 2465  
QY 2385 TAATATAAAGAGATATGTTCACAGTTCCAGTTCCCAACAGCCAAATACCAGATGTGCATTTCT 2444  
Db 2466 CACCTCCCCAGCTGAACITTTTCCACCTGGAGTCCCTTGGGAATACCGGACGTGATCTCTT 2525  
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAAATGSCCGATCAACTGCTGTGAAAAAT 2504  
Db 2526 TTATAGTCCAATGATGTGACCCAGTCTCTGCAGTTCTGSGAGATCAACCAACCATCCGCGT 2585  
QY 2505 GAGGTATACTCTACTAAATCTGGAGCAGGAGTGATTTTCACTCCCCAGCAAGTGCCGACG 2564  
Db 2586 CAGGTGCAGTCCACAGAAAACTGTCCCTGGAGTTTGTGCTGCCAGGAACGTGCTCAGA 2645  
QY 2565 AGTACCTGTGATGGGTGTACGTTCTATTTCCTGTGGGAGAGTGTGAAGCTTGCCCTCT 2624  
Db 2646 TGGGACCTGTGATGGCTGCAACTTCCACTTCCACTTCTGTGGGAGAGCGCGGCTGTGCCCGCT 2705  
QY 2625 GTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCTGCAAGAGAGGATTTTCAGGAAC 2684  
Db 2706 CTGCTCAGTGGCTGACTACCATGCTATCGTCAGCAGCTGTGTGGTGGGATCCAGANGAC 2765  
QY 2685 CTTGTATGTGGAATGAACCTTAAATGGTGCATTAAGGAAATTTCTTTGCCCTGAGAAAAA 2744  
Db 2766 TACTTACGTGTGNCGAGAACCCCAAGCTATGCTCTGTGGCATTTTCTCTGCTGAGCAGAG 2825  
QY 2745 GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGGTGGGAGCGGTGTGGGAGCTTT 2804  
Db 2826 AGTCACCATCTGCAAAACCATAGATTCTTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG 2885  
QY 2805 TACTGCCGTTTGTGGTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATCAAAAACCTGGA 2864  
Db 2886 TACTGCCATCTCTGCTCACCGCTTTTGACCTGCTACTTTTGGAAAAAGAAATCAAAAACCTAGA 2945  
QY 2865 ATACAAATATTCGAAGTTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCCGGCTGC 2924  
Db 2946 GTACAAGTACTCCAAGCTGGTGATGAATGCTACTCTCAAGGACTGTGACCTGCCAGCAGC 3005  
QY 2925 AGACAGTTGTGCTATCATGGAAGGAGAGAATAATGAAGAGGAAGTTGTATATTTCCAATAA 2984  
Db 3006 TGACAGCTGCGCCATCATGGAAGCGGAGGATGTAGAGGACGACCTCATCTTTACCAGCAA 3065

```

QY  2985  ACAGTCACTACTAGGAAAACTCAAATCTTTGGCAACCCAGGAAAAAGAGACCATTTTGA  3044
Db      ||||| | | | | ||||| | | | ||||| | | | ||||| | | | |||||
3066  GAAGTCACTTTTGGGAAGATCAAATCATTACCTCCAAGAGGACTCCTGATGGATTGA  3125
QY  3045  ATCTGTTCAACTGAAACCTC  3065
Db      || | | ||||| || | |
3126  CTCAGTGGCGCTGAAGACATC  3146

RESULT 15
US-10-176-921-37
; Sequence 37, Application US/10176921
; Publication No. US2003002726A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C288
; CURRENT APPLICATION NUMBER: US/10/176,921
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 37
; LENGTH: 3501
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 2762, 2778
; OTHER INFORMATION: unknown base
US-10-176-921-37

```

QY	471	GGGCCCTTCTGACAGCAGGCCAGACGGCTGTAAACAACCTCTTCTTGGATCCCTCGTGGAAA	530
Db	567	CAGTGTCTGAGTCCA--CCGGAACTGTACTTCGTCCAAGTGGGTTCCTCCCGGGCGGA	623
QY	531	CTACATAGAAATCTAATCGTGTATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAA	590
Db	624	CTACATCGCCTCCAAACACGGACGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA	683
QY	591	GAAGTCAGGCTATGTCTTCTTTGAGTACCAGTATGTTCGACAAACAACATCTTCTTTTGAGTTT	650
Db	684	GCAATCTGGCACCGTTAACTTCGAAATCTACTATCCAGACTCCAGCATCATCTTTTGAGTTT	743
QY	651	CTTTATTCAAATGATCAGTGCACGGAGATGGACACACCACCTGACAAAGTGGTAAACCT	710
Db	744	TTTCGTTTCAGAAATGACCAAGTGCACGCCAAATGCAGATGACATC--CAGTGGATGAAGAC	800
QY	711	TACAGACAATGGAGAAATGGGGCTCTCATTTCTGTAATGCTGAATCAGGCACAAACATAT	770
Db	801	CACAGAGAA--AGGATGGGAATTCACACAGTGTGGAGCTAATCGAGGCAATAATGTCTT	857
QY	771	CTACTGGAGAACTACAGGGCACTTATGGGTTCTAAGGCGGTCAAGCCGTGTGCTGGTAAA	830
Db	858	CTATTGGAGAACACAGCCCTTCTCAGTATGGACCAAAAGTACCCAAGCCTGTGCTGGTGAG	917
QY	831	AAATATACAATTGAAGGGGTGGCTACACATCAGAAATGTTTTCCTTGCAGGCCAGGCAC	890
Db	918	AAACAATTGCCATAACAGGGGTGGCTACACTTCAGAAATGCTTCCCCCTGCAAAACCTGGCAC	977
QY	891	ATTGAGCAACAACACAGGTTTCATTCAACTGCCAGGTGTGCCAGAAACACACTATTTCTGA	950
Db	978	GTATGCAGACAAGCAGGGCTCCTCTTCTGCAAACTTTGCCACGCCAACTCTTATTCAAA	1037
QY	951	GAAAGGAGCCAAAGAATGTATAAGGTGTAA--AGACGACTCTCAATTTTCAGGATCCAG	1007
Db	1038	TAAAGGAGAAACTTCTTGCCACGAGTGTGACCCCTGACAAATCTCAGAGAAAGGATCTTC	1097
QY	1008	TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAGACTATTTCCAGATCCATCTCCATG	1067
Db	1098	TTCTCTTAACGTGCGCCAGCTTGCACAGACAAAGATATTTCTACACACACACGGCCTG	1157
QY	1068	TGATGAAGAAGGAAGACACAGATAAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGGA	1127
Db	1158	CGATGCCAACGGAGAGACACAAACTCATGTACAAATGGGCCANGCCGAAATCTGTAGCGA	1217
QY	1128	GGATCTCACAGATGCTATTAGATTGCCCCCTTCTCGAGAGAGAAGGATTTGCCGCCCTTG	1187
Db	1218	GGACCTTGAGGGGGCAGTGAAGCTGCCTGCCCTCTGGTGTGAAGACCCACTGCCACCCCTG	1277
QY	1188	CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTCTGGAAACAT	1247
Db	1278	CAACCCAGGCTTCTTCAAAACCAACAACAGCACTGCCAGCCCTGCCCATATGGTTCTCTA	1337
QY	1248	TTTCAGATGGAAACCAAGAATGTAGACCAATGTCCAGCAGGAACGGAGCTTGCCTT	1307
Db	1338	CTCCAATGG--CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGAT	1394
QY	1308	TGAATATAATGGTGGAAATGTCTTCTTGGCAACATGAAAACCTTCTTCAATGTATGG	1367
Db	1395	TGAATACAAATGGTGGAAACACGCTGCCCAACAACATGGAACGACCCGTTCTCAGTGGGAT	1454
QY	1368	GAATTCAAAGTGGATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGG	1427
Db	1455	CAACTTCGAGTACAAGGGCATGACAGGCTGGGAGGTGGCTGGTGTACATTTACACACG	1514
QY	1428	GGCTGGAGGTCTTGACAAATGATTACCTTGATCTTAACTTGCATATCCCAGGATTTAAACC	1487
Db	1515	TGCTGGAGCCTTCACACAATGACTTCATGATTTCTCACTCTGGTTGTGCCAGGATTTAGACC	1574
QY	1488	ACCAACATCTATGACTGGAGCCAC--GGGTTCTGAACCTAGGAGAATAACATTTGTCTTT	1544
Db	1575	TCCGCGAGTCTGATGGCAGACACAGAGAAATAAGAGGTGGCCAGAAATCAACATTTGTCTT	1634
QY	1545	TGAGACCCCTCTGTTTCAGCTGACTGTGTTTGTACTTTCATGGTGGATATTAATAGAAAAAG	1604

Db 1635 TGAGACCTCTGTTCTGTGAACCTGTGAGCTCTACTTTCATGGTGGGTGTGAATTCCTAGGAC 1694  
QY 1605 TACAAATGTGGTAGAATCGTGGGTGGAACCAAGAAACCAAGCTTACACCCATATCAT 1664  
Db 1695 CAACACTCCTGTGGAGACGTGGAAGGTTCCAAAGGCAACAGTCTTATACCTATACATCAT 1754  
QY 1665 CTTCAGAATGCAACTTTTACATTTACATGGGCATTCAGAGAACTAATCAGGTCGAAGA 1724  
Db 1755 TGAGGAAACACTACACAGAGCTTCACCTGGGCTTCCAGAGGACCACTTTTTCATGAGGC 1814  
QY 1725 TAATAGACGGTTCATCAATGACATGGTGAAGATTTATCTATCACAGCCACTAATGCAGT 1784  
Db 1815 AAGCAGGAAGTACACCAATGACGTGCCAAGATCTACTCCATCAATGTCACCAATGTAT 1874  
QY 1785 TGATGGGTGGCGTCCCTCATGCCGTGCCCTGTGCCCTCGGTTCTGAACAGTCGGGTTTCATC 1844  
Db 1875 GAATGGGTGGCCTCCTACTGCCGTCCCTGTGCCCTAGAACCTCTGATGTGGGCTCCTC 1934  
QY 1845 GTGTGTCCCCTGCCCTCCAGGCCACTACATTTGAGAAAGAAACCAACCAAGTGCAGGAATG 1904  
Db 1935 CTGCACCTCTGTCTGTGTTACTATATTGACCGAGATTACAGGAACCTGCCACTCCTG 1994  
QY 1905 TCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGATTTCCATG 1964  
Db 1995 CCCCCCTAACACAATTTCTGAAAGCCCAACCAAGCTTATGGTGTCAGGCTGTGTGCCCTG 2054  
QY 1965 CGGGCCTGGAGTAAAAACAATCAGACCACTTCGGTTTGCTATAGTACTGCTTTTCTA 2024  
Db 2055 TGGTCCAGGACCAAGAACAACAGATCCACTCTCTGTGCTACAATGATGCACTTCTC 2114  
QY 2025 CCATGAAAAAGAAAATCAGATTTTGCACTATGACTTTAGCAACCTCAGCAGTGTGGGCTC 2084  
Db 2115 ACGCAACACTCCAACCAAGGACTTTCAACTACAACCTTCTCCGCTTTGGCAACACCCGTCAC 2174  
QY 2085 ATTAATGAATGSCCCCGAGCTTCACCTCCAAAGGAACAAAATACTTCCATTTCTTCAATAT 2144  
Db 2175 TCTTGCTGGAGSGCCCAAGCTTCACTTCCAAGGGTTGAAATACTTCCATCACTTTACCCCT 2234  
QY 2145 CAGTTTATGTGGSCATGAGGGGAAGAAGATGGCTCTCTGTACCACAAATATAACAGACTT 2204  
Db 2235 CAGTCTCTGTGGAACCAAGGAGGTAGGAARATGTCTGTGTGCACCCGACAATGTCAGTACCT 2294  
QY 2205 TACAGTAAAGAAATAGTGGCAGGGTCAGATGATTACACAAAATTTGGTAGGGGCATTTGT 2264  
Db 2295 CCGGATTCCTGAGGGTGAGTCAGGG-----TTCTCCAAATCTATCACAGCCTACGT 2345  
QY 2265 ATGCCAGTCAACAATATTATCTCTGAAAGTAAGGGTTTCCGAGCAGCCTTATCATCAC 2324  
Db 2346 CTGCCAGGCAGTCATCATCCCCCAGAGGTGACAGGCTACAAGGCCGGGGTTTCCTCACA 2405  
QY 2325 ATCCATCATTTCTGGCAGATACATTCATAGGATFCACAGTTGAAACCACATTTGAAAAATAT 2384  
Db 2406 GCCTGTCAGCCTTGCTGATCGACTTATTGGGCTGACAACAGATATGACTCTGGATGGAAT 2465  
QY 2385 TAATATAAAGAAGATATGTTCCCAGTTCCACAAAGCCAAAATACCAGATGTGCAATTTCTT 2444  
Db 2466 CACCTCCCCAGCTGAACCTTTCCACCTGGAGTCCCTTGGGAATACCGGACGTGATCTTCTT 2525  
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAATGGCCGATCAACTGCTGTGAAAAAT 2504  
Db 2526 TTATAGGTCCAATGATGACCCAGTCCCTGCAGTTCTGGGAGATCAACCACCACTCCGCGT 2585  
QY 2505 GAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCCCAGCAAGTGGCCAGC 2564  
Db 2586 CAGGTGCAGTCCACAGAAAACTGTCCCTGGAAGTTTGCTGCTGCCAGGACCTGCTCAGA 2645  
QY 2565 AGGTACCTGTGATGGGTGACGTTCTATTTCCTGTGGGAGAGTGCTGAAGCTTGGCCCTCT 2624  
Db 2646 TGGACCTGTGATGGCTGCAACTTCCACTTCCCTGTGGGAGAGCGCGCTGTGCCCCGCT 2705  
QY 2625 GTGTACGGAGCATGACTTCCATGAGATTTGAGGGAGCCTGCAAGAGAGGATTTTCAGGAAAC 2684

Db 2706 CTGCTCAGTGGCTGACTACCATGCTATCGTCAGCAGCTGTGTGGTGGGATCCAGANGAC 2765  
QY 2685 CTTGTATGTGTGAATGAACCTAAATGGTGCATTAAGGAATTTCTTTGCCCTGAGAAAAA 2744  
Db 2766 TACTTACGTGTGNCAGAAACCCAAAGCTATGCTCTGGTGGCATTTCTCTGCTGAGCAG 2825  
QY 2745 GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGGTGGGAGCCGCTGTGGGAGCTTT 2804  
Db 2826 AGTCACCATCTGCAAAACCATAGATTTCTGCTGAAAGTGGGCATCTCTGCAGGCACCTG 2885  
QY 2805 TACTGCCGTTTTTGGTGGTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATCAAAAACTGGA 2864  
Db 2886 TACTGCCATCCTGCTCACCGTCTTGACCTGCTACTTTTGGAAAAAGAAATCAAAAACTAGA 2945  
QY 2865 ATACAAATATTCCAAGTTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCCCGGCTGC 2924  
Db 2946 GTACAAAGTACTCCAAGCTGGTGATGAATGCTACTCTCAAGGACTGTGACCTGCCAGCAGC 3005  
QY 2925 AGACAGTTGTGCTATCATGGAAGGAGAAGATATGAAGAGGAAGTTGTATATTTCCAATAA 2984  
Db 3006 TGACAGCTGCGCCATCATGGAAGCGGAGGATGTAGAGGACGACCTCATCTTTACCAGCAA 3065  
QY 2985 ACAGTCACTACTAGGAAAACTCAAAATCTTTGGCAACCAAGAAAAAGAGACCATTTTGA 3044  
Db 3066 GAAGTCACTTTTGGGAAGATCAAAATCATTTACCTCCAAGAGGACTCCTGATGGATTGA 3125  
QY 3045 ATCTGTTCAACTGAAAACTC 3065  
Db 3126 CTCAGTGGCGCTGAAGACATC 3146

Search completed: May 12, 2003, 13:07:44  
Job time : 566.762 secs





score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 00:01:37 ; Search time 8400.33 Seconds  
(without alignments)  
11744.602 Million cell updates/sec

Title: US-10-073-333A-1  
Perfect score: 3390  
Sequence: 1 atgtctgttcgcgcgcgggg.....ggtataaaaaaaaaaaaaa 3390

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

SUMMARIES

Result NO.	Score	Match	Length	DB	ID	Description
1	2617.2	77.2	3185	9	AK055902	AK055902 Homo sapi
2	1499.6	44.2	1737	6	AX060311	AX060311 Sequence
3	1182.6	34.9	1508	6	AX060301	AX060301 Sequence
4	904.6	26.7	3280	6	AX036667	AX036667 Sequence
5	903	26.6	3331	6	AX127726	AX127726 Sequence
6	901	26.6	3501	6	AX463904	AX463904 Sequence
7	860.8	25.4	2610	6	AX127742	AX127742 Sequence
8	860.8	25.4	2733	6	AX127740	AX127740 Sequence
9	709.6	20.9	2895	9	AK057647	AK057647 Homo sapi
10	600.2	17.7	2449	9	BC031648	BC031648 Homo sapi
11	547.	16.1	2129	9	AK026832	AK026832 Homo sapi
12	517.6	15.3	136150	9	AC002081	AC002081 Homo sapi
13	517.2	15.3	5567.	9	AB037745	AB037745 Homo sapi
14	493.2	14.5	1129	6	AR184096	AR184096 Sequence
15	482.2	14.2	576	6	AX336856	AX336856 Sequence
16	400.4	11.8	2188	10	BC029010	BC029010 Mus muscu
17	400.2	11.8	1717	6	AX013071	AX013071 Sequence
18	381.8	11.3	1149	6	AX127746	AX127746 Sequence
19	285.8	8.4	1119	6	AX213279	AX213279 Sequence
20	285.8	8.4	1587	6	AX213277	AX213277 Sequence
21	265.4	7.8	101719	2	AL357495	AL357495 Homo sapi
22	234.6	6.9	627	6	AX127744	AX127744 Sequence
23	217	6.4	131321	2	AC080160	AC080160 Mus muscu
24	198	5.8	404	6	AX071721	AX071721 Sequence
25	196.4	5.8	101719	2	AL357495	AL357495 Homo sapi
26	196.2	5.8	1149	10	BC022655	BC022655 Mus muscu
27	192.8	5.7	165865	2	AC079958	AC079958 Mus muscu
28	189	5.6	107889	9	AC004519	AC004519 Homo sapi
29	180.4	5.3	108063	2	AC116297	AC116297 Rattus no
30	161.6	4.8	426	6	AX368203	AX368203 Sequence
31	140.6	4.1	437	6	AX182039	AX182039 Sequence
32	135.4	4.0	480	6	AX182004	AX182004 Sequence
33	130.4	3.8	29979	9	AC000122	AC000122 Homo sapi
34	127.2	3.8	131321	2	AC080160	AC080160 Mus muscu
35	123.6	3.6	353	6	AX182022	AX182022 Sequence
36	118.6	3.5	166384	2	AC113446	AC113446 Mus muscu
37	118	3.5	108063	2	AC116297	AC116297 Rattus no
38	115	3.4	399	6	AX182000	AX182000 Sequence
39	114	3.4	183815	2	AC099081	AC099081 Rattus no
40	112.6	3.3	150129	2	AC119789	AC119789 Rattus no
41	104	3.1	443	6	AX127734	AX127734 Sequence
42	99.8	2.9	97114	9	AL356389	AL356389 Human DNA
43	94.4	2.8	677	6	AX036669	AX036669 Sequence
44	94	2.8	265	6	AX368154	AX368154 Sequence
45	90.8	2.7	89129	10	AL672200	AL672200 Mouse DNA

ALIGNMENTS

RESULT 1  
AK055902  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE

AK055902 3185 bp mRNA linear PRI 01-AUG-2002  
Homo sapiens cDNA FLJ31340 fis, clone MESAN1000035, weakly similar  
to MAJOR SURFACE-LABELLED TROPHOZOITE ANTIGEN PRECURSOR.

AK055902  
AK055902.1 GI:16550745  
oligo capping; fis (full insert sequence).  
Homo sapiens normal mesangial cells (NHMC56046-2) cDNA to mRNA,  
clone\_lib:MESAN1 clone:MESAN1000035.

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS  
1 Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,

Fukuzumi,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Irie,R.,  
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,  
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,  
Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,  
Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,  
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,  
Nagahari,K., Masuo,Y., Nagai,K. and Isogai,T.  
NEDO human cDNA sequencing project  
Unpublished

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

2 (bases 1 to 3185)  
Isogai,T., Otsuki,T. and Sugiyama,T.  
Direct Submission  
Submitted (24-Oct-2001)

Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: RAB and  
HRI.

FEATURES

source

Location/Qualifiers  
1..3185  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MESAN1000035"  
/cell\_type="normal mesangial cells (NHMC56046-2)"  
/clone\_lib="MESAN1"  
/note="cloning vector: pME18SFL3-primary culture, normal  
mesangial cells"  
362..1843

CDS

/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB71041.1"  
/db\_xref="GI:16550746"  
/translation="MLEGNYQVMLKSGTNILYWRRTGILMGSKAVKPVLVKNITIEGV  
AYTSECFPKPGTFSNKPGEFNCQVCPNRTYSEKAKECIRCKDDSQFSEESSECTE  
RPGCTTKDYEQIHTPCDEGKTQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNP  
GFYNGSSSCHPCPPGTFSDGTEKCRPCPAGTEPALGFYKWNVLPGNMKTSCEFNVG  
NSKCDGMNGWEVAGDHIOGAGGSDNDYLILNLHIPGFKPPTSMGTGAGSELGRITFV  
FETLCSADCVLYFMVDINRKTNNVESGGTKEQAYTHIEFNATFTTWFQRTNQ  
GQDNRREINDMVKIYSITATNAVGVASSCRACALGSEQSSGSCVPCPPGHYIEKETN  
QCKECPDPTYLSIHQVYKEACIPCGPGSKNNODHSVCYSDCFFYPEKENQSLHYDFS  
NLSSVGLMNGPSTSKTKYFHFNFNISYVGMRRRWLSVPTI"  
BASE COUNT 960 a 660 c 719 g 846 t

Query Match 77.2%; Score 2617.2; DB 9; Length 3185;  
Best Local Similarity 93.2%; Pred. No. 0;  
Matches 2866; Conservative 0; Mismatches 3; Indels 207; Gaps 4;

QY	504	CAACTCTTCTGGATCCCTCGTGGAACTACATAAGATCTAATCGTGATGACTGCACG	563
Db	111	CAGCTCTTCTGGATCCCTCGTGGAACTACATAAGATCTAATCGTGATGACTGCACG	170
QY	564	GTCTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGCTCTTTGAGTACCAG	623
Db	171	GTCTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGCTCTTTGAGTACCAG	230
QY	624	TGTCGACAAACATCTTCTTTGAGTCTTTATTCAAATGATCATGCTGCCAGGAGATG	683
Db	231	TGTCGACAAACATCTTCTTTGAGTCTTTATTCAAATGATCATGCTGCCAGGAGATG	290
QY	684	CACCACCACTGACAAGTGGGTAAACTTACAGACAATGGAGAATGGGGCTCTCATCT	741
Db	291	CACCACCACTGACAAGTGGGTAAACTTACAGACAATGGAGAATGGGGCTCTCATCT	350
QY	742	-----GTAATGCTGAAATCAGGCAACATAC	769
Db	351	CAGCTCCTTAAATGCTTGAGGCAATTACCAGGTAATGCTGAAATCAGGCAACATAC	410

QY	770	TCTACTGGAGAACTACAGGCATCCTTTATGGGTCTTAAGCGGTCAGCCTGTGCTGGTAA	829
Db	411	TCTACTGGAGAACTACAGGCATCCTTTATGGGTCTTAAGCGGTCAGCCTGTGCTGGTAA	470
QY	830	AAATATCACAAATTGAAGGGGTGGGTACACATCAGAATGTTTCTTCAAGCCAGGCA	889
Db	471	AAATATCACAAATTGAAGGGGTGGGTACACATCAGAATGTTTCTTCAAGCCAGGCA	530
QY	890	CATTTCAGCAACAAACCAGGTTTCATTCAACTGCCAGGTGTGCCAGAAACACCTATTCTG	949
Db	531	CATTTCAGCAACAAACCAGGTTTCATTCAACTGCCAGGTGTGCCAGAAACACCTATTCTG	590
QY	950	AGAAAGGAGCCAAAGAATGTATAAGGTGTAAAGACGACTCTCAATTTTC-----AGGAT	1003
Db	591	AGAAAGGAGCCAAAGAATGTATAAGGTGTAAAGACGACTCTCAATTTTCAGAGGAAGGAT	650
QY	1004	CCAGTGAGGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATATC	1063
Db	651	CCAGTGAGGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATATC	710
QY	1064	CATGTGATGAAGAAAGGAAAGACACAGATATGTACAAGTGGATAGAGCCCAAAATCTGCC	1123
Db	711	CATGTGATGAAGAAAGGAAAGACACAGATATGTACAAGTGGATAGAGCCCAAAATCTGCC	770
QY	1124	GGGAGGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAAAGGATTTGTCGC	1183
Db	771	GGGAGGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAAAGGATTTGTCGC	830
QY	1184	CTTGCAACCCCTGGATTTTATACAAATGGATCATCTTCTTGCCATCCCTGTCTCTGGAA	1243
Db	831	CTTGCAACCCCTGGATTTTATACAAATGGATCATCTTCTTGCCATCCCTGTCTCTGGAA	890
QY	1244	CATTTTCAGATGGAACCAAGAAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCCTTG	1303
Db	891	CATTTTCAGATGGAACCAAGAAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCCTTG	950
QY	1304	GCTTTGAATATAAATGGTGGATGTCTTCTTGGCAACATGAAAACCTTCTGCTCAATG	1363
Db	951	GCTTTGAATATAAATGGTGGATGTCTTCTTGGCAACATGAAAACCTTCTGCTCAATG	1010
QY	1364	TTGGGAATTCAAAGTCCGATGGAATGAATGGTTGGAGGTGGCTGGAGATCATATCCAGA	1423
Db	1011	TTGGGAATTCAAAGTCCGATGGAATGAATGGTTGGAGGTGGCTGGAGATCATATCCAGA	1070
QY	1424	GTGGGGCTGGAGGTTCTGACAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTA	1483
Db	1071	GTGGGGCTGGAGGTTCTGACAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTA	1130
QY	1484	AACCAACCAACATCTATGACTGGAGCCACGGGTTCTGAACATAGGAAGAAATAACATTTGTCT	1543
Db	1131	AACCAACCAACATCTATGACTGGAGCCACGGGTTCTGAACATAGGAAGAAATAACATTTGTCT	1190
QY	1544	TTGAGACCCCTCTGTTCAGCTGACTGTGTTTGTACTTTCATGGTGGATATTATAGAAAAA	1603
Db	1191	TTGAGACCCCTCTGTTCAGCTGACTGTGTTTGTACTTTCATGGTGGATATTATAGAAAAA	1250
QY	1604	GTACAAATGTGGTAGAATCGTGGGGTGGAAACCAAGAAACAAAGCTTACACCCATATCA	1663
Db	1251	GTACAAATGTGGTAGAATCGTGGGGTGGAAACCAAGAAACAAAGCTTACACCCATATCA	1310
QY	1664	TCTTCAAGAATGCAACTTTTACATTTACATGGGCATTTCCAGAGAATTAATCAGGGTCAAG	1723
Db	1311	TCTTCAAGAATGCAACTTTTACATTTACATGGGCATTTCCAGAGAATTAATCAGGGTCAAG	1370
QY	1724	ATAATAGACGGTTCATCAATGACATGGTGAAGATTTATTTCTATCACAGCCACTAATGCAG	1783
Db	1371	ATAATAGACGGTTCATCAATGACATGGTGAAGATTTATTTCTATCACAGCCACTAATGCAG	1430
QY	1784	TTGATGGGGTGGGCTCTCATGCCGTGCCCTGTGCCCTCGGTTCTGAACAGTCGGGTTCAT	1843
Db	1431	TTGATGGGGTGGGCTCTCATGCCGTGCCCTGTGCCCTCGGTTCTGAACAGTCGGGTTCAT	1490
QY	1844	CGTGTGTCCCTTSCCTCCAGGCCACTACATTTGAGAAAGAAACCAACCAAGGAAAT	1903



Db 1491 ||||| CGTGTGCCCTGCCCTCCAGGCCACTACATTGAGAAAGAAACCAACCAAGTGCAAGGAAT 1550

QY 1904 GTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGATTCCAT 1963

Db 1551 GTCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGATTCCAT 1610

QY 1964 GCGGGCTGGGAGTAAAAACAATCAGGACCATTCCGGTTTGCTATAGTGACTGCTTTTCT 2023

Db 1611 GCGGGCTGGGAGTAAAAACAATCAGGACCATTCCGGTTTGCTATAGTGACTGCTTTTCT 1670

QY 2024 ACCATGAAAAAGAAATCAGATTTTGCACATATGACTTTAGCAACCTCAGCAGTGTGGGCT 2083

Db 1671 ACCCTGAAAAAGAAATCAGAGTTTGCACATATGACTTTAGCAACCTCAGCAGTGTGGGCT 1730

QY 2084 CATTAAATGAATGGCCCCAGCTTACCTCCAAAGGAACAAAATACCTTCCATTCTTCAATA 2143

Db 1731 CATTAAATGAATGGCCCCAGCTTACCTCCAAAGGAACAAAATACCTTCCATTCTTCAATA 1790

QY 2144 TCAGTTTATGTGGGCATGAGGGGAAGAGATGGCTCTCTGTACCAACAATATAACAGACT 2203

Db 1791 TCAG-TTATGTGGGCATGAGGGGAAGAGATGGCTCTCTGTACCAACAATATAACAGACT 1849

QY 2204 TTACAGTAAAAAGAAATAGTGGCAGGTCAGATGATTACACAAATTTGGTAGGGGCATTTG 2263

Db 1850 TTACAGTAAAAAGAAATAGTGGCAGGTCAGATGATTACACAAATTTGGTAGGGGCATTTG 1909

QY 2264 TATGCCAGTCAACAATTATTCCCTCTTGAAAGTAAGGGTTTCCGAGCAGCCTTATCATCAC 2323

Db 1910 TATGCCAGTCAACAATTATTCCCTCTTGAAAGTAAGGGTTTCCGAGCAGCCTTATCATCAC 1969

QY 2324 AATCATCATCTTGGCAGATACATTATAGGAGTCACAGTGAACACCACATTTGAAAAATA 2383

Db 1970 AATCATCATCTTGGCAGATACATTATAGGAGTCACAGTGAACACCACATTTGAAAAATA 2029

QY 2384 TTAATATAAAAGAGATATGTTCCAGTTCCAAAGCCAAATACCAGATGTGCATTTCT 2443

Db 2030 TTAATATAAAAGAGATATGTTCCAGTTCCAAAGCCAAATACCAGATGTGCATTTCT 2089

QY 2444 TTTATAAGTCTTCTACAGCAACAACATCTTGTTATTAATGGCCGATCAACTGCTGTGAAAA 2503

Db 2090 TTTATAAGTCTTCTACAGCAACAACATCTTGTTATTAATGGCCGATCAACTGCTGTGAAAA 2149

QY 2504 TGAGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTCAGTCCCCAGCAAGTGCCCCAG 2563

Db 2150 TGAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTCAGTCCCCAGCAAGTGCCCCAG 2209

QY 2564 CAGGTACCTGTGATGGGTGTACGTTCTATTTCCTGTGGGAGAGTGCTGAAGCTTGCCCTC 2623

Db 2210 CAGGTACCTGTGATGGGTGTACGTTCTATTTCCTGTGGGAGAGTGCTGAAGCTTGCCCTC 2269

QY 2624 TGTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCTGCAAGAGAGGATTTTCAGGAAA 2683

Db 2270 TGTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCTGCAAGAGAGGATTTTCAGGAAA 2329

QY 2684 CCTGTATGTGTGGAATGAACCTAAATGGTGCATTAAAGGAATTTCTTTGCCTGAGAAAA 2743

Db 2330 CCTGTATGTGTGGAATGAACCTAAATGGTGCATTAAAGGAATTTCTTTGCCTGAGAAAA 2389

QY 2744 AGTTGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGGTGGGAGCCGGTGTGGGAGCTT 2803

Db 2390 AGTTGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGGTGGGAGCCGGTGTGGGAGCTT 2449

QY 2804 TTACTGCCGTTTGTGCTGGTGGCTCTGACCTGTACTTCTGAAAAAGAAATCAAAA----- 2858

Db 2450 TTACTGCCGTTTGTGCTGGTGGCTCTGACCTGTACTTCTGAAAAAGAAATCAAAAACCTGG 2509

QY 2859 ----- 2858

Db 2510 AATACAAATATTCCAAGTTAGTAATGACGACTAACTCAAAAAGAGTGTGAACCTCCCGGCTG 2569

QY 2859 ----- 2858

Db 2570 CAGACAGTTGTGTATCATGGAAGGAGAAGATAATGAAGAGGAAGTTGTATATTCCAATA 2629

QY 2859 -----GAAAAAGAAAGACCATTTTG 2877

Db 2630 AACAGTCACTACTAGGAAAACTCAAAATCTTTGGCAACCAAGGAAAAAGAAAGACCATTTTG 2689

QY 2878 AATCTGTCAAACTGAAAAACCTCAAGATCCCCAAATATATGAAGAGACAGTGTGTAGCCT 2937

Db 2690 AATCTGTCAAACTGAAAAACCTCAAGATCCCCAAATATATGAAGAGACAGTGTGTAGCCT 2749

QY 2938 TGAGACTAATGAACAAAGAAACCTGCTCTAGTTTACAGGACCATAATTTAGGGTCTGTC 2997

Db 2750 TGAGACTAATGAACAAAGAAACCTGCTCTAGTTTACAGGACCATAATTTAGGGTCTGTC 2809

QY 2998 CTCATACCTGTCACATTGGTGATCTCACAGAGAGGGGCCATGCCGCTGAAAAAGGAAGGA 3057

Db 2810 CTCATACCTGTCACATTGGTGATCTCACAGAGAGGGGCCATGCCGCTGAAAAAGGAAGGA 2869

QY 3058 GATTGAAACATTTGATTGCCCTTATCAGATGGTCAAGTACCTTGCCAAATAAAGGAAAGCA 3117

Db 2870 GATTGAAACATTTGATTGCCCTTATCAGATGGTCAAGTACCTTGCCAAATAAAGGAAAGCA 2929

QY 3118 AATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAAGAGATTTATCTGTATATA 3177

Db 2930 AATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAAGAGATTTATCTGTATATA 2989

QY 3178 CACATAACTGAAAAACCAAGTTTAAGCCCCACCAATGCACCTGCTGATGCCATATAATT 3237

Db 2990 CACATAACTGAAAAACCAAGTTTAAGCCCCACCAATGCACCTGCTGATGCCATATAATT 3049

QY 3238 AATGGGTAACTTTTATTCTTTATGATGTCTACATAACAAGTGTGATTTGGAGGCACATG 3297

Db 3050 AATGGGTAACTTTTATTCTTTATGATGTCTACATAACAAGTGTGATTTGGAGGCACATG 3109

QY 3298 TGAGCATATGCATTATGATCCAATTTATGTTTTTCTTTTGTATATTGGGAAAAATT 3357

Db 3110 TGAGCATATGCATTATGATCCAATTTATGTTTTTCTTTTGTATATTGGGAAAAATT 3169

QY 3358 AAAATTTTTTAAGGT 3373

Db 3170 AAAATTTTTTAAGGT 3185

RESULT 2

AX060311

LOCUS AX060311 1737 bp DNA linear PAT 22-JAN-2001

DEFINITION Sequence 19 from Patent WO0078802.

ACCESSION AX060311

VERSION AX060311.1 GI:12405800

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1737)

AUTHORS Shimkets,R.A., Fernandes,E., Vernet,C., Yang,M., Boldog,F.L. and Herrmann,J.L.

TITLE Secreted polypeptides and corresponding polynucleotides

JOURNAL Patent: WO 0078802-A 19 28-DEC-2000;

Curagen Corporation (US)

FEATURES

Location/Qualifiers

1..1737

/organism="Homo sapiens"

/db\_xref="taxon:9606"

296..1690

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAC24875.1"

/db\_xref="GI:12405801"

/translation="MKNQVCCKGEGTYSLGSGIKFDEWDELPAQFSNIATFMDTVVG

PSDSRPDGCNNSSWIIPRGNYIESNRDDCTVSLIYAVHLKKSQYVFFEYQYVDNFI

FFIQNDQCQEMDFTTDKWKVLTLDNGEWGSHSVMLKSKSNILYWRTTGILMGSKAVKPV

LVKNITIEGVAYTSECPCKPGTFSNKPQSFNCQVCPRNTYSEKGAKECIRCKDDSQF

CDS

SEEGSSECTERPPCTTKDYFQIHPTPCDEEGKTQIMYKWKIEPKICREDLTDAILRPPSG  
EKDKPCPNPGFYNNGSSSCHPCPPGTFSDGTEKCRPGAGTEPALGFYKWNVLPG  
NMKTSFCNFVNGSKCDGMNGWEVAGDHIQSGAGSDNDYLILNLHIPGFKPPTSMTGAT  
GSELGRITFVFETLCSADCVLVFMVDINRKSTNVESWGGTKEKOAYTHIIFKNATFT  
FTWGIPIRELIOGPR"

BASE COUNT 495 a 381 c 409 g 451 t 1 others  
ORIGIN

Query Match 44.2%; Score 1499.6; DB 6; Length 1737;  
Best Local Similarity 98.7%; Pred. No. 0;  
Matches 1545; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

QY	189	GAAGATTATCACCTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC	248
Db	142	GAAGATTATCACCTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC	201
QY	249	CATTCCAAAATCTGCAGTGGACTGCTCTGGCCTGCCTGACCCAGTGAGAGGCAAAAGATG	308
Db	202	CATTCCAAAATCTGCAGTGGACTGCTCTGGCCTGCCTGACCCAGTGAGAGGCAAAAGATG	261
QY	309	CACCTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGCAGTAAGTG	368
Db	262	CACCTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGCAGTAAGTG	321
QY	369	TGGTGAAGGCACCTATTCCCTTGGCAGTGGCATCAAATTTGATGAATGGGATGAATTGCC	428
Db	322	TGGTGAAGGCACCTATTCCCTTGGCAGTGGCATCAAATTTGATGAATGGGATGAATTGCC	381
QY	429	GGCAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGCACAGCAG	488
Db	382	GGCAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGCACAGCAG	441
QY	489	GCCAGACGGCTGTAACAACCTCTTCTTGGATCCCTCGTGGAACACTACATAGAATCTAATCG	548
Db	442	GCCAGACGGCTGTAACAACCTCTTCTTGGATCCCTCGTGGAACACTACATAGAATCTAATCG	501
QY	549	TGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAAGAAAGTCAGGCTATGTCTT	608
Db	502	TGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAAGAAAGTCAGGCTATGTCTT	561
QY	609	CTTTGAGTACCAGTATGTCGACAAACATCTTCTTTGAGTCTTTATTCAAAAATGATCA	668
Db	562	CTTTGAGTACCAGTATGTCGACAAACATCTTCTTTGAGTCTTTATTCAAAAATGATCA	621
QY	669	GTGCCAGGAGATGGACACCACCACCTGACAAGTGGGTAAACCTTACAGACAATGGAGAATG	728
Db	622	GTGCCAGGAGATGGACACCACCACCTGACAAGTGGGTAAACCTTACAGACAATGGAGAATG	681
QY	729	GGGCTCTCATCTGTAAATGCTGAATCAGGCACAAACATACCTCTACTGGAGAACTACAGG	788
Db	682	GGGCTCTCATCTGTAAATGCTGAATCAGGCACAAACATACCTCTACTGGAGAACTACAGG	741
QY	789	CATCCTTATGGGTTCTAAGCGGTCAGGCCTGTGCTGGTAAAAAATATCACAATTTGAAG	848
Db	742	CATCCTTATGGGTTCTAAGCGGTCAGGCCTGTGCTGGTAAAAAATATCACAATTTGAAG	801
QY	849	GGTGGGTACACATCAGAAATGTTTCTTGAAGCCAGGCACATTCAGCAACAACCCAGG	908
Db	802	GGTGGGTACACATCAGAAATGTTTCTTGAAGCCAGGCACATTCAGCAACAACCCAGG	861
QY	909	TTCATTCACTGCCAGGTGTGTCGCCAACAACACCTATTCTGAGAAAGGAGCCAAAGAATG	968
Db	862	TTCATTCACTGCCAGGTGTGTCGCCAACAACACCTATTCTGAGAAAGGAGCCAAAGAATG	921
QY	969	TATAAGGTGTAAAGACGACTCTCAATTTTC-----AGGATCCAGTGAGTGTACAGAGCG	1022
Db	922	TATAAGGTGTAAAGACGACTCTCAATTTTCAGAGGAAGGATCCAGTGAGTGTACAGAGCG	981
QY	1023	CCCTCCCTGTACCACAAAAGACTATTTCAGATCCATCTCCATCTGATGAAGAAGGAAA	1082
Db	982	CCCTCCCTGTACCACAAAAGACTATTTCAGATCCATCTCCATCTGATGAAGAAGGAAA	1041
QY	1083	GACACAGATAATGTACAAGTGGATAGAGCCCCAAAATCTGCCGGGAGGATCTCACAGATGC	1142

Db	1042	GACACAGATAATGTACAAAGTGGATAGAGCCCCAAAATCTGCCGGGAGGATCTCACAGATGC	1101
QY	1143	TATTAGATTGCCCCCTTCTGGAGAGAAAGAGATTGTCCGCCCTTGCAACCCTGGATTTA	1202
Db	1102	TATTAGATTGCCCCCTTCTGGAGAGAAAGAGATTGTCCGCCCTTGCAACCCTGGATTTA	1161
QY	1203	TAACAATGGATCATCTTCTGGCATCCCTGTCTCCTCGGAAACATTTTCAGATGGAAACCA	1262
Db	1162	TAACAATGGATCATCTTCTGGCATCCCTGTCTCCTCGGAAACATTTTCAGATGGAAACCA	1221
QY	1263	AGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGCTTTGAATATAAATGGTG	1322
Db	1222	AGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGCTTTGAATATAAATGGTG	1281
QY	1323	GAATGTCTCTTCCCTGGCAACATGAAACCTTCCCTGCTTCAATGTGGGAATTCAAAAGTCCGA	1382
Db	1282	GAATGTCTCTTCCCTGGCAACATGAAACCTTCCCTGCTTCAATGTGGGAATTCAAAAGTCCGA	1341
QY	1383	TGGAATGAATGGTTGGGAGGTGGTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGA	1442
Db	1342	TGGAATGAATGGTTGGGAGGTGGTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGA	1401
QY	1443	CAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATGAC	1502
Db	1402	CAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCACCAACATCTATGAC	1461
QY	1503	TGGAGCCACGGGTTCTGAACCTAGGAAGAAATAACATTTGTCTTTGAGACCCCTCTGTTTCAGC	1562
Db	1462	TGGAGCCACGGGTTCTGAACCTAGGAAGAAATAACATTTGTCTTTGAGACCCCTCTGTTTCAGC	1521
QY	1563	TGACTGTGTTTGTACTTTCATGTTGGATATTAATAGAAAAAGTACAAAATGTGGTAGAATC	1622
Db	1522	TGACTGTGTTTGTACTTTCATGTTGGATATTAATAGAAAAAGTACAAAATGTGGTAGAATC	1581
QY	1623	GTGGGTGGAACCAACAAAGAAACAAAGCTTACACCCATATCATCTTCAAGAATGCAACCTTT	1682
Db	1582	GTGGGTGGAACCAACAAAGAAACAAAGCTTACACCCATATCATCTTCAAGAATGCAACCTTT	1641
QY	1683	TACATTTACATGGG--CATTCAGAGAACTAAT--CAGGGTCAAGATAATAGACGGTTCA	1738
Db	1642	TACATTTACATGGGGCATTTCCAGAGAACTAATTCAGGGTCCAAGATAATAGACGGTTCC	1701
QY	1739	TCAAT 1743	
Db	1702	NCCAT 1706	

RESULT 3  
AX060301  
LOCUS AX060301 1508 bp DNA linear PAT 22-JAN-2001  
DEFINITION Sequence 9 from Patent WO0078802.  
ACCESSION AX060301  
VERSION AX060301.1 GI:12405790  
KEYWORDS human.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1. (bases 1 to 1508)  
AUTHORS Shimkets,R.A., Fernandes,E., Vernet,C., Yang,M., Boldog,F.L. and Herrmann,J.L.  
TITLE Secreted polypeptides and corresponding polynucleotides  
JOURNAL Patent: WO 0078802-A 9 28-DEC-2000;  
FEATURES  
source  
1. .1508  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
226. .1461  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC24870.1"

/db\_xref="GI:12405791"  
/translation="MHSSWIPRGNVIESNRDDCTVSLIYAVHLKSGYVFFEYQYVDN  
NIFFEFFIQQCOEMDTTDDKWKLTJDNGEWSHVMKSGTNILYWRTTGILMSGK  
AVKPVLVKNITIEGVAYTSECFCKPGTFENKPGSFNCQVCPRNTYSEKGAKECIRCK  
DDSFSEGSSECTERPPCTTKDYFIHTPCDEEGKQIIMYKWIPEKICREDLTDAIR  
LPPSGEKDKDPPCNPFGFYNNSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFEYKWW  
NVLPGNMKTSFCFNVGNSKCDGMNGWEVAGDHIOQSGAGGSDNDYLILNLHIPGFKPPTS  
MTGATGSELGRITFVFETLCSADCVLVFMVDINRKSTNVVESWGGTKEKQAYTHIIFK  
NATFTFTWGPRIEIQGR"

BASE COUNT	444 a	306 c	355 g	402 t	1 others
ORIGIN					
Query Match	34.9%;	Score 1182.6;	DB 6;	Length 1508;	
Best Local Similarity	98.4%;	Pred. No. 3.1e-266;			
Matches 1228;	Conservative 0;	Mismatches 10;	Indels 10;	Gaps 3;	
QY	506	ACTCTTTCTGGATCCCTCGTGGAAACTACATAGAAATCTAATCGTGATGACTGCGACGGTGT	565		
Db	230	ACTCTTCTTGGATCCCTCGTGGAAACTACATAGAAATCTAATCGTGATGACTGCGACGGTGT	289		
QY	566	CTTTGATCTATGCTGTGCACCTTAAGAAAGTCAGGCTATGTCTTTTGGATACCAATG	625		
Db	290	CTTTGATCTATGCTGTGCACCTTAAGAAAGTCAGGCTATGTCTTTTGGATACCAATG	349		
QY	626	TCGACAACAACATCTTCTTTGAGTTCTTTATTCAAAATGATCAGTGCCAGGAGATGGACA	685		
Db	350	TCGACAACAACATCTTCTTTGAGTTCTTTATTCAAAATGATCAGTGCCAGGAGATGGACA	409		
QY	686	CCACCACACTGACAAGTGGGTAAAACTTACAGACAATGGAGAATGGGGCTCTCATTTCTGTAA	745		
Db	410	CCACCACACTGACAAGTGGGTAAAACTTACAGACAATGGAGAATGGGGCTCTCATTTCTGTAA	469		
QY	746	TGCTGAAATCAGGCACAAACATCTACTTGAGAACTACAGGCATCCTTTATGGGTCTA	805		
Db	470	TGCTGAAATCAGGCACAAACATCTACTTGAGAACTACAGGCATCCTTTATGGGTCTA	529		
QY	806	AGGCGGTCAAGCCTGTGCTGGTAAAAAATATCACAAATTGAAAGGGTGGCGTACACATCAG	865		
Db	530	AGGCGGTCAAGCCTGTGCTGGTAAAAAATATCACAAATTGAAAGGGTGGCGTACACATCAG	589		
QY	866	AATGTTTTCCTTGCAAGCCAGGCACATTCAGCAACAAACCAGGTTTCAATCAACTGCCAGG	925		
Db	590	AATGTTTTCCTTGCAAGCCAGGCACATTCAGCAACAAACCAGGTTTCAATCAACTGCCAGG	649		
QY	926	TGTGTCAGAAAAACACCTATTCTGAGAAAGGAGCCAAAGAAATGTATAAGGTGTAAAGACG	985		
Db	650	TGTGTCAGAAAAACACCTATTCTGAGAAAGGAGCCAAAGAAATGTATAAGGTGTAAAGACG	709		
QY	986	ACTCTCAATTTTC-----AGGATCCAGTGAGTGTACAGAGCGCCCTCCCTGTACCACAA	1039		
Db	710	ACTCTCAATTTTCAGAGGAAGGATCCAGTGAGTGTACAGAGCGCCCTCCCTGTACCACAA	769		
QY	1040	AAGACTATTTCCAGATCCATACTCCATGTGATGAAGAAGGAAAGACACACAGATAATGTACA	1099		
Db	770	AAGACTATTTCCAGATCCATACTCCATGTGATGAAGAAGGAAAGACACACAGATAATGTACA	829		
QY	1100	AGTGGATAGAGCCCCAAAATCTGCCGGGAGGATCTCACAGATGCTATTAGATTGCCCCCTT	1159		
Db	830	AGTGGATAGAGCCCCAAAATCTGCCGGGAGGATCTCACAGATGCTATTAGATTGCCCCCTT	889		
QY	1160	CTGGAGAGAAGAGGATGTGCCGCCCTTGCAACCCCTGGATTTTATAACAATGGATCATCTT	1219		
Db	890	CTGGAGAGAAGAGGATGTGCCGCCCTTGCAACCCCTGGATTTTATAACAATGGATCATCTT	949		
QY	1220	CTTGCCATCCCTGTCTCCTGGAACATTTTCAGATGGAACCAACAAATGTAGACCATGTC	1279		
Db	950	CTTGCCATCCCTGTCTCCTGGAACATTTTCAGATGGAACCAACAAATGTAGACCATGTC	1009		
QY	1280	CAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAATGGTGGAAATGTCCTTCCTGGCA	1339		
Db	1010	CAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAATGGTGGAAATGTCCTTCCTGGCA	1069		
QY	1340	ACATGAAAACCTCTCTGCTTCAATGTGTTGGGAATTCAAAGTGGCATGGAATGAATGGTTGG	1399		

Db	1070	ACATGAAAACCTTCCTGCTTCAATGTTGGGAATTCAAAAGTCGGATGGAATGAATGGTTGGG	1129		
QY	1400	AGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGACAATGATTACCTGATCT	1459		
Db	1130	AGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGACAATGATTACCTGATCT	1189		
QY	1460	TAAACTTGCAATATCCAGGATTTAAACCCACCAACATCTATGACTGGAGCCACGGTTCTG	1519		
Db	1190	TAAACTTGCAATATCCAGGATTTAAACCCACCAACATCTATGACTGGAGCCACGGTTCTG	1249		
QY	1520	AACTAGGAAGAATAACATTTGCTTTTGAGACCCCTCTGTTTCTGACTGACTGTGTTGTACT	1579		
Db	1250	AACTAGGAAGAATAACATTTGCTTTTGAGACCCCTCTGTTTCTGACTGACTGTGTTGTACT	1309		
QY	1580	TCATGGTGGATATTAAATAGAAAAAGTACAAAATGTGTAGAAATCGTGGGTGGAAACCAAAG	1639		
Db	1310	TCATGGTGGATATTAAATAGAAAAAGTACAAAATGTGTAGAAATCGTGGGTGGAAACCAAAG	1369		
QY	1640	AAAAACAAGCTTACACCCATATCATCTTCAAGAATGCAAACTTTTACATTTTACATGGG--C	1697		
Db	1370	AAAAACAAGCTTACACCCATATCATCTTCAAGAATGCAAACTTTTACATTTTACATGGGCA	1429		
QY	1698	ATTCCAGAGAACTAAT--CAGGGTCAAGATAATAGACGGTTTCAATCAAT	1743		
Db	1430	TTCCCAGAGAACTAATTCAGGGTCCAAGATAATAGACGGTTCCNCCAT	1477		
RESULT 4					
AX036667					
LOCUS	AX036667	Sequence 1 from Patent WO0058460.	3280 bp	DNA	linear
DEFINITION	AX036667				
ACCESSION	AX036667				
VERSION	AX036667.1	GI:11226242			
KEYWORDS		human.			
SOURCE		human.			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
TITLE		1 (bases 1 to 3280)			
JOURNAL		Bruck,C.E., Vinals,Y.D., Coche,T. and Cassart,J.P.			
FEATURES		Novel compounds			
source		Patent: WO 0058460-A 1 05-OCT-2000;			
		BRUCK CLAUDINE ELVIRE MARIE (BE) ; SMITHKLINE BEECHAM BIOLOG (BE) ;			
		VINALS Y DE BASSOLS CARLOTA (BE) ; COCHE THIERRY (BE) ; CASSART			
		JEAN POL (BE)			
		Location/Qualifiers			
		1..3280			
		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			
BASE COUNT	810 a	930 c	797 g	743 t	
ORIGIN					
Query Match	26.7%;	Score 904.6;	DB 6;	Length 3280;	
Best Local Similarity	60.1%;	Pred. No. 5.4e-201;			
Matches 1637;	Conservative 0;	Mismatches 1059;	Indels 27;	Gaps 7;	
QY	171	GCTTCCTCCTTGCCAGGAGAAAGATTATCACCTTTGAATATACGGAATGTATAGCAGTGG	230		
Db	135	GCTTCATGCCTGCAAGAGCTGTGACTACCACCTATGAGTACACGGCGTGTGACAGCACGGG	194		
QY	231	CTCCAGGTGGAGAGTTGCCATTCCAAATTTGCGAGTGGAGTGTCTTGGCCCTGCCTGACCC	290		
Db	195	TTCCAGGTGGAGGGTGGCCGTGCCGCATACCCCGGCCCTGTGCACCAGCCTGCCTGACCC	254		
QY	291	AGTGAGAGGCCAAAGAAATGCACCTTTCTCCTGTGCTTCTTGGAGAGTATCTAGAAATGAAGAA	350		
Db	255	CGTCAAGGGCACCCGAGTGTCTCTCTCTGCAACGCCGGGAGTTTCTGGATATGAAGGA	314		
QY	351	CCAGGTATGCAGTAAGTGTGGTGAAGGCACCTATTCTTGGCAGTGGCATCAAATTTGA	410		
Db	315	CCAGTCAATGAAGCCATGCGCTGAGGGCCGCTACTCCCTCGGCACAGGCATTCGGTTTGA	374		



QY 411 TGAATGGGATGAATTGCCGGCAGGATTTTCTAACATCGCAACATTTCAATGGACACTGTGGT 470  
Db 375 TGAGTGGGATGAGCTGCCCCATGGCTTTGCCAGCCTCTCAGCCAAACATGGAGCTGGATGA 434  
QY 471 GGGCCCTTCTGACAGCAGGCCAGACGGCTGTAACAACTCTTCTTGGATCCCTCGTGGAAA 530  
Db 435 CAGTGTCTGAGTCCA---CCGGGAACGTGACTTCGTCCAAGTGGGTTCGCCGGGGCGA 491  
QY 531 CTACATAGAATCTAATCGTGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAA 590  
Db 492 CTACATCGCCTCCAACACGGACGCAATGCACGCCACACTGATGTACGCCGTCAACCTGAA 551  
QY 591 GAAGTCAGGCTATGTCTTCTTTGAGTACCAGTATGTCGACAACAACATCTTCTTTGAGTT 650  
Db 552 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 611  
QY 651 CTTTATTCAAAAATGATCAGTGCCAGGAGATGGACACCACCACACTGACAAGTGGGTAAAAC 710  
Db 612 TTTCTGTTCAGAATGACCAGTGCCAGGCCCAATGCAGATGACTC---CAGGTGGATGAAGAC 668  
QY 711 TACAGACAATGGAGAAATGGGGCTCTCATTTCTGTAAATGCTGAAATCAGGCACAACATACT 770  
Db 669 CACAGAGAA---AGGATGGGAATTCACAGACTGTGGAGCTAAATCGAGGCAATAATGTCT 725  
QY 771 CTACTGGAGAACTACAGGCATCCTTATGGGTCTTAAGGGCTGTAAGGGCTCAAGCCTGTGCTGTA 830  
Db 726 CTATTGGAGAACACAGCGCTTCTCAGTATGGACCAAAAGTACCCAAAGCCTGTGCTGGTGAG 785  
QY 831 AAATATCACAAATGAAGGGGTGGCGTACACATCAGAATGTTTTCCTTGCAAGCCAGGCAC 890  
Db 786 AAACATTGCCATAACAGGGGTGGCCTACACTTCAGAATGCTTCCCTGCAAACTTGGCAC 845  
QY 891 ATTCAGCAACAAACAGGTTTCATTCAACTGCCAGGTGTGTCCCGAAGAACACCTATTCTGA 950  
Db 846 GTATGCAGACAAGCAGGGCTCCTCTTCTGCAAACTTTTGCCAGCAACTCTTATTCAA 905  
QY 951 GAAAGGAGCCAAAGAATGTATAAGGTGFAAA---GACGACTCTCAATTTTCAGGATCCAG 1007  
Db 906 TAAAGGAGAACTTCTTGCCACCAGTGTGACCCTTGACAAATACTCAGAGAAAGGATCTTC 965  
QY 1008 TGAGTGTACAGCGCCCTCCCTGTACACAAAAGACTATTTCCAGATCCATCTCCATG 1067  
Db 966 TTCTGTAACTGCGCCAGCTTGCACAGACAAAGATTATTTCTACACACACACGGCCTG 1025  
QY 1068 TGATGAAGAAAGAAAGACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGGA 1127  
Db 1026 CGATGCCAACGGAGAGACAACTCATGTACAAATGGGCCAAGCCGAAATCTGTAGCGA 1085  
QY 1128 GGATCTCACAGATGCTATTAGATTGGCCCTTCTGGAGAGAGAAGGATTTGTCGGCCTTG 1187  
Db 1086 GGACCTTGAGGGGCAGTGAAGCTGCCTGCTCTGGTGTGAAGACCCACTGCCACCCCTG 1145  
QY 1188 CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCCTGGAACAT 1247  
Db 1146 CAACCCAGGCTTCTTCAAAACCAACACAGCACCTGCCAGCCCTGCCCATATGGTTCCTA 1205  
QY 1248 TTCAGATGGAACCAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTT 1307  
Db 1206 CTCCAATGG---CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGAT 1262  
QY 1308 TGAATATAATGGTGAATGTCTTCTTGCAACATGAAACTTCTGCTTCAATGTTGG 1367  
Db 1263 TGAATACAATGGTGAACACGCTGCCCAACAACATGGAACACCGCTTCTCAGTGGGAT 1322  
QY 1368 GAATTCAAAGTCGATGGAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427  
Db 1323 CAACCTCGAGTACAAGGGCATGACAGGCTGGGAGGTGGCTGGTGTATCACATTTACACAGC 1382  
QY 1428 GGCTGGAGGTTCTGACAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACC 1487  
Db 1383 TGCTGGAGCCTCAGACAATGACTTCATGATTCTCACTCTGGTGTGCCAGGATTTAGACC 1442  
QY 1488 ACCAACATCTATGACTGGAGCCAC---GGGTTCTGAACTAGGAAGAAATAACATTTGTCTT 1544

Db 1443 TCCGCAGTCGGTGGATGGCAGACACAGAGAATAAAGAGGTGGCCAGAATCACATTTGTCIT 1502  
QY 1545 TGAGACCCCTCTGTTCAGCTGACTGTGTTTGTACTTTCATGGTGGATATTAATAGAAAAAG 1604  
Db 1503 TGAGACCCCTCTGTCTGTGAACGTGTGAGCTCTACTTTCATGGTGGGTGTAATTCTAGGAC 1562  
QY 1605 TACAAATGTGGTAGAATCGTGGGTGGAAACCAAGAAACAAAGCTTACACCCATATCAT 1664  
Db 1563 CAACACTCCTGTGGAGACGTGGAAGGTTCCAAAGGCAACAGTCTCTATACCTACATCAT 1622  
QY 1665 CTTCAAGAATGCAACTTTTACATTTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGA 1724  
Db 1623 TGAGGAGAACACTACCACGAGCTTCACCTGGGCTTCCAGAGGACCACTTTTCATGAGGC 1682  
QY 1725 TAATAGACGGTTTCATCAATGACATGGTGAAGATTTTATCTATCACAGCCACTAATGCAGT 1784  
Db 1683 AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCAACCAATGTTAT 1742  
QY 1785 TGATGGSGTGGCGTCCCTCATGCCCCTGCTGCCCCCTCGGTTCTGAACAGTCCGGTTCTATC 1844  
Db 1743 GAATGGCGTGGCCTCCTACTGCCCCTGCTGCCCCCTAGAAAGCCTCTGATGTGGGCTCCTC 1802  
QY 1845 GTGTGTCCCCCTGCCCTCCAGGCCACTACATTTGAGAAAGAAACCAACAGTGCAGAGGAATG 1904  
Db 1803 CTGCACCTCTTGTCTCTGCTGGTTACTATATTTGACCGAGATTTCAGGAACCTGCCACTCCTG 1862  
QY 1905 TCCACCTGACACCTACCTGTCTCATACATCAGGTCTATGGCAAGAGGCTTGTATTCATG 1964  
Db 1863 CCCCCCTAACACAATCTGAAAGCCCCACAGCCTTATGTTGTCCAGGCTGTGTGCCCTG 1922  
QY 1965 CGGCGCTGGGAGTAAAAACAATCAGAGCCATTCGGTTTGTATAGTACTGCTTCTTCTA 2024  
Db 1923 TGGTCCAGGGACCAAGAACACAAGATCCACTCTCTGTGTACAAATGATGCACCTTCTC 1982  
QY 2025 CCATGAAAAAGAAATCAGATTTTGCACTATGACTTTAGCAACCTCAGCAGTGTGGGCTC 2084  
Db 1983 ACGCAACACTCCAACCCAGGACTTTCAACTACAACCTTCTCCGCTTGGCAACACACCCGTCAC 2042  
QY 2085 ATTAATGAATGGCCCCAGCTTCACCTCCAAGGAACAAAATACCTTCCATTTCTTCAATAT 2144  
Db 2043 TCTTGTGTGGAGGGCCAAGCTTCACCTCCAAGGGTTGAAATACTTCCATCACCTTACCCCT 2102  
QY 2145 CAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTT 2204  
Db 2103 CAGTCTCTGTGGAAACCAAGGTAGGAANAATGTCTGTGTGCACCCGACAATGTCACTGACCT 2162  
QY 2205 TACAGTAAAAGAAATAGTGGCAGGGTCAGATGATTACACAANAATTTGGTAGGGGCAATTTG 2264  
Db 2163 CCGGATTCCTGAGGGTGAGTCAGGG-----TTCCTCCAATCTATCACAGCCTACCT 2213  
QY 2265 ATGCCAGTCAACAATTAATTCCTTCTGAAAAGTAAGGGTTTCCGAGCAGCCTTATCATCACAA 2324  
Db 2214 CTGCCAGGCAGTCATCATCCCCCAGAGGTGACAGGCTACAAAGGCCGGGTTCCTCACAA 2273  
QY 2325 ATCCATCATTTCTGGCAGATTACATTCATAGGAGTCACAGTTGAAACCACATTTGAAAAATAT 2384  
Db 2274 GCCTGTCAAGCCTTGTGTGATCGACTTATTTGGGGTGACAACAGATATGACTCTGGATGGAA 2333  
QY 2385 TAATATAAAGAAGATATGTTCCCAAGTTCCAACAAGCAAAATACCAAGATGTGCATTTCTT 2444  
Db 2334 CAGCTCCCAGCTGAACCTTTCCACCTGGAGTCCCTTGGGAATACCGGACGTGATCTTCTT 2393  
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAATGGCCGATCAACTGCTGTGAAAAAT 2504  
Db 2394 TTATAGGTCCAATGATGTGACCCAGTCTCTGCAGTTCTGGGAGATCAACACCACCTCCCGCT 2453  
QY 2505 GAGGTGTAATCCTACTAATCTGGAGCAGGAGTGTATTCAGTCCCCCAGCAAGTCCCCAGC 2564  
Db 2454 CAGGTGCAGTCCACAGAAAAACTGTCCCTGGAAAGTTTGTCTGCTGCCAGGAACGTGCTCAGA 2513  
QY 2565 AGGTACCTGTGTGATGGGTGTACGTTCTATTTCTGTGGGAGAGTGTGTAAGCTTGGCCTCT 2624

Db 2514 TGGGACCTGTGATGGCTGCAACTTCCACTTCCTGTGGGAGAGCGGCTGTGCTTGGCCGCT 2573

Qy 2625 GTGTACGGAGCATGACTTCCATGAGATTGAGGAGCCTGCAAGAGAGAGGATTTCAGGAAAC 2684

Db 2574 CTGCTAGTGGCTGACTACCATGCTATCGTCAGCAGCTGTGTGGTGGGATCCAGAGAC 2633

Qy 2685 CTTGTATGTGTGAATGAACCTAAATGGTGCATTAAAGGAATTCTTTGCCCTGAGAAAAA 2744

Db 2634 TACTTACGTGTGGGAGAGACCAAGCTATGCTCTGGTGGCATTTCTCTGCCTGAGCAGAG 2693

Qy 2745 GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGGTGGGAGCCGGTGTGGGAGCTTT 2804

Db 2694 AGTCACCATCTGCAAAACCATAGATTCTTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG 2753

Qy 2805 TACTGCGGTTTGTGGTGGCTGTGACCTGCTACTTCTGGAAGAAAGAAATCAAAAGAAAA 2864

Db 2754 TACTGCCATCTGCTCACCGTCTTGACCTGCTACTTTTGGAAAAAGAAATCAAAACTAGA 2813

Qy 2865 GAAGACCATTTTGAATCTGTTCA 2887

Db 2814 GTACAAGTACTCCAAGCTGGTGA 2836

RESULT 5

AX127726

LOCUS AX127726 3331 bp DNA linear PAT 15-MAY-2001

DEFINITION Sequence 1 from Patent WO0131003.

ACCESSION AX127726

VERSION AX127726.1 GI:14134373

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 3331)

TITLE Delneste,Y., Magistrelli,G., Jeannin,P. and Bonnefoy,J.Y.

JOURNAL Cloning, expression and characterisation of a gene expressed in tumour cells and involved in the regulation of the immune response

PATENT: WO 0131003-A 1 03-MAY-2001;

PIERRE FABRE MEDICAMENT (FR)

FEATURES

Location/Qualifiers

1..3331

/organism="Homo sapiens"

/db\_xref="taxon:9606"

45..3086

CDS

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAC38907.1"

/db\_xref="GI:14134374"

/translation="MAEPGSHLSARVGRTERIPRLWRLLLWAGTAFQVTQGTGP ELHACKESEHYHEYTACDSTGSRWRVAVPHTPGICTSLDPVKGTGECFSCNAGEFLD MKDQCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAESTGCTSSKWV PRGDYIASNTDECTATLMYAVNLKQSGTVNFYYPDSSIIEFFVQNDQCPNADDS RWMKTEKGWEFHSVELNRGNVLYWRTTAFSVYTKVPKPLVLRNIAITGVAYTSECF PCKPGYADKQSSFCCLCPANSYSNKGTSCHQCDPKYSEKSSSCNVRPACTDKD YFYHTACDANGETOLMYKWAKPKICSEDLGAVKLPASGVKTHCPCNPGFFKTNNS TCQPCYPGYSNGSDTRCPAGTEPAVGFEYKWNWTLPTNMETTVLSGINFEYKGMTG WEVAGDHIYTAAGASDNDFMILTIVVPGFRPPPOSMADTENKEVARITEFEETLCSVN CELYFMVGVNSRNTNPVETWKSGKQSYTYIIEENTTSFTWAFQRTTPEASRKYT NDVAKIYSINVTNVMNGVASYCRPCALEASDVSSCTSCPAGYIDRDSGTCHSCPNN TILKAHQPYGVQACVPCPGPTKNNKIHSLCYNDCFTSRNTPTRTFNFNFSALANTVTL AGPSFTSKGLYFHFHTLSLCGNQGRKMSVCTDNVTLRIPEGESGFSKSIAYVCQ AVIIPPEVTGYKAGVSQPVSLADRLIGVTTMDTLDGITSPAELFHLSESLGIPDVIFE YRSNDVTQSCSSGRSTIRVRCSPQKTVPGSLLPGTCSDGTCDCGNFHLWESAAC PLCSVDYRAIVSSCVAGIQKTYVWREPKLCSGSIPLQQRVTICKTIDFWLKVGIS AGTCTAILTLVLCYFWKKNQKLEYKYSKLVNMATLKDCLDLPAAADSCAIMEGEDVEDD LIFTSKKSLEFGIKSFKRTPDGFDSVPLKTSGGPMDL"

BASE COUNT 822 a 945 c 813 g 751 t

ORIGIN

Query Match 26.6%; Score 903; DB 6; Length 3331;

Best Local Similarity 60.1%; Pred. No: 1.3e-200;

Matches 1636; Conservative 0; Mismatches 1060; Indels 27; Gaps 7;

Qy 171 GCTTCCTCCTTGCCAGGAGAAAGATTATCACCTTTGAATATACGGAATGTGATAGCAGTGG 230

Db 179 GCTTCATGCCTGCAAAAGAGTCTGAGTACCACCTATGAGTACACGGCGTGTGACAGCACGGG 238

Qy 231 CTCCAGGTGGAGAGTTGCCATTCCAAATTTCTGCAGTGGAGCTGCTGTGGCCTGCCTGACCC 290

Db 239 TTCCAGGTGGAGGGTGGCGTGGCGCATACCCCGGGCCTGTGACACAGCCTGCCTGACCC 298

Qy 291 AGTGAGAGGCAAAAGAAATGCACCTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAA 350

Db 299 CGTCAAGGGCACCGAGTGTCTCTCTGCAACGCCGGGAGTTTCTGGATATGAAGGA 358

Qy 351..CCAGTATGCAGTAAGTGTGGTGAAGGCACCTATTCTTGGGAGTGGCATCAAAATTTGA 410

Db 359 CCAGTCATGTAAGCCATGCGCTGAGGGCCGCTACTCCCTCGGCACAGGCATTCGGTTTGA 418

Qy 411 TGAATGGGATGAATTGCCGGCAGGATTTCTAAACATCGCAACATTCATGGACACTGTGGT 470

Db 419 TGAGTGGGATGAGTGCCTCCCATGGCTTTGCCAGCCTCTCAGCAACATGGAGCTGGATGA 478

Qy 471 GGGCCCTTCTGACAGCAGGCGGCAGCGGCTGTAAACAACCTCTTCTGGATCCCTCGTGGAAA 530

Db 479 CAGTCTGCTGAGTCCA--CCGGGAAGTGTACTTCTGTCCTCAAGTGGGTCCCCGGGGCGGA 535

Qy 531 CTACATAGAAATCTAATCGTGTGATGACTGCACGGTGTCTTTTGATCTATGCTGTGCACCTTAA 590

Db 536 CTACATCGCCTCCAACACGGACGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 595

Qy 591 GAAGTCAGGCTATGTCTTCTTCAAGTACCAGTATGTCGACAACAACATCTTCTTTGAGTT 650

Db 596 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 655

Qy 651 CTTTATTCAAAATGATCAGTGCAGGAGATGGACACCACCTGACAAAGTGGGTAAAACT 710

Db 656 TTTTCGTTCAGAATGACCAGTGCACGCCCAATGCAGATGACTC--CAGGTGGATGAAGAC 712

Qy 711 TACAGACAATGGAGAATGGGGCTCTCATTTCTGTAATGCTGAATCAGGCACAAACATACT 770

Db 713 CACAGAGAA--AGGATGGGAATTCACACAGTGTGGAGCTAAATCGAGGCAATAATGTCTCT 769

Qy 771 CTACTGGAGAACTACAGGCATCCTTATGGGTTCTAAGGGCGGTCAAGCCCTGTGCTGGTAAA 830

Db 770 CTATTGGAGAACCAACAGCCTTCTCAGTATGGACCACCAAGTACCCAAAGCCTGTGCTGGTGA 829

Qy 831 AAATATCACAAATTGAAGGGTGGCGTACACATCAGAAATGTTTCTTCTGCAAGCCAGGCAC 890

Db 830 AAACATTGCCATAACAGGGGTGGCGCTACACTTCAGAAATGCTTCCCTGCAAAACCTGGCAC 889

Qy 891 ATTACAGCAACAAACCAGGTTTCATTCAACTGCCAGGTGTGTCCAGAAACACCTATTCTGA 950

Db 890 GTATGCAGACAAAGCAGGGCTCCTCTTCTGTCAAACCTTTGCCAGCAACCTCTTATTCAA 949

Qy 951 GAAAGGAGCCAAAGAAATGTATAAGGTGTAATA--GACGACTCTCAATTTTCAGGATCCAG 1007

Db 950 TAAAGGAGAAACTTCTTGCCACCAGTGTGACCCCTGACAAATACTCAGAGAAAGGATCTTC 1009

Qy 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCACTCCCATG 1067

Db 1010 TTCCTGTAACTGCGCCCGAGCTTGACAGACAAAGATTATTTACACACACACAGGCCTG 1069

Qy 1068 TGATGAAGAAGAAAGACACAGATATGTACAAGTGGATAGAGCCCAAAATCTGCCGGGA 1127

Db 1070 CGATGCCAACGGAGAGACACAACTCATGTACAAATGGGCCAAGCCGCAAAATCTGTAGCGA 1129

Qy 1128 GGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAAAGAGATTGTCCGCCCTG 1187

Db 1130 GGACCTTGAGGGGGCAGTGAAGCTGCTGCTCTGTTGTGAAGACCCACTGCCACCCCTG 1189

Qy 1188 CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTCTGGAACATT 1247

Db 1190 CAACCCCGGCTTCTTCAAAACCAACACAGCACCTGCCAGCCCTGCCCATATGTTCCCTA 1249



QY	1248	TTCAGATGGAACCAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCCTGCACCTTGGCTT	1307
Db	1250	CTCCAATGG--CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGAT	1306
QY	1308	TGAATATAAATGGTGAATGTCTCTCCCTGGCAACATGAAAACCTCCTGCTTCAATGTTGG	1367
Db	1307	TGAATACAAATGGTGAACACACGCTGCCCACAAACATGGAACGACCCGTTCTCAGTGGGAT	1366
QY	1368	GAATTCAAAGTCGGATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGG	1427
Db	1367	CAACTTCGAGTACAAGGCATGACAGGCTGGGAGGTGGCTGCTGATCACATTTACACAGC	1426
QY	1428	GGCTGGAGGTCTTGACAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAACC	1487
Db	1427	TGCTGGAGCCTCAGACAATGACTTCATGATTTCTCACTCTGGTTGTGCCAGGATTTAGACC	1486
QY	1488	ACCAACATCTATGACTGGAGCCAC---GGGTTCTGAACCTAGGAAGAATAACATTTGTCTT	1544
Db	1487	TCCGCAGTCGGTGATGGCAGACACAGAGATAAAGAGGTGGCCAGAGATCACATTTGTCTT	1546
QY	1545	TGAGACCCCTCTGTTCAGCTGACTGTGTTTGTACTTTCATGGTGGATATTAAAGAAAAAG	1604
Db	1547	TGAGACCCCTCTGTCTGTGAACCTGTGAGCTCTACTTTCATGGTGGGTGTGAATTTCTAGGAC	1606
QY	1605	TACAAATGTGGTAGAATCGTGGGTGGAAACCAAGAAAAACAAGCTTACACCCATATCAT	1664
Db	1607	CAACACTCTGTGGAGACGTGGAAGGTTCCAAAGGCCAAACAGCTCTATACCTACATCAT	1666
QY	1665	CTTCAAGAAATGCAACTTTTACATTTACATGGGCAATTCAGAGAACTAATCAGGTTCAAGA	1724
Db	1667	TGAGGAGAACACTACCACGAGCTTCACCTGGGCCTTCCAGAGGACCACTTTTCATGAGGC	1726
QY	1725	TAATAGACGGTTCATCAATGACATGGTGAAGATTTATTCTATACAGCCCACTAATGCAGT	1784
Db	1727	AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCACCAATGTTAT	1786
QY	1785	TGATGGGGTGGCGTCCCTCATGCCGTGCCGTGCTGCCCTCGGTTCTGAACAGTCGGGTTCATC	1844
Db	1787	GAATGGCGTGGCCTCCTACTGCCGTCCCTGTGCCCTAGAAAGCCTCTGATGTGGGCTCCTC	1846
QY	1845	GTGTGTCCCTGCCCTCCAGGCCACTACATTTGAGAAAGAAACCAACCAGTGCAGGAATG	1904
Db	1847	CTGCACCTCTGTCTCCTGTGTTACTATATTGACCGAGATTCAGGAACCTGCCACTCCTG	1906
QY	1905	TCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGAGGCTTGATTCCATG	1964
Db	1907	CCCCCTTAACACAATTCTGAAAGCCCAACGACCTTATGGTGTCCAGGCCTGTGTGCCCTG	1966
QY	1965	CGGGCCTGGGAGTAAACAAACATCAGGACCACTTCGGTTTGCTATAGTGACTGCTTTTCTA	2024
Db	1967	TGGTCCAGGGACCAAGAACAACAAGATCCACTCTCTGTGTGCTACAAATGATTCACCTTCTC	2026
QY	2025	CCATGAAAAAGAAAATCAGATTTTGCATGACTTTTAGCAACCTCAGCAGTGTGGGCTC	2084
Db	2027	ACGCAACACTCCAACACAGGACTTTCAACTACAACCTTCTCCGCTTTGGCAACACCCGTCA	2086
QY	2085	ATTAATGAATGGCCCCAGCTTCACCTCCAAGGAACAAAATACTTCCATTTCTTCAATAT	2144
Db	2087	TCTTGCTGGAGGCCCAAGCTTCACCTCCAAGGGTTGAAATACTTCCATCACTTTACCCCT	2146
QY	2145	CAGTTTATGTGGSCATGAGGGGAAGAAGATGGCTCTCTGTACCACCAATATAACAGACTT	2204
Db	2147	CAGTCTCTGTGGAACACCAAGGTAGGAAATGTCTGTGTGCACCCGACAATGTCACGTGACCT	2206
QY	2205	TACAGTAAAGAAATAGTGGCAGGGTCAGATGATTACACAAAATTTGGTAGGGGCAATTTGT	2264
Db	2207	CCGGATTCCTGAGGGTGAGTCAGGG-----TTCTCCAAATCTATCACAGCCCTACGT	2257
QY	2265	ATGCCAGTCAACAATATTCTTCTGAAGTAAGGGTTTCCGAGAGCCCTTATCATCACAC	2324
Db	2258	CTGCCAGGCAGTCATCATCCCCCAGAGGTGACAGGCTACAAGGCCGGGGTTTCTCTCAC	2317
QY	2325	ATCCATCATTTCTGGCAGATACATTTCATAGGAGTCACAGTTTGAAACCAACCAATTGAAAAATAT	2384

Db	2318	GCCTGTCAGCCTTGCTGATCGACTTATTGGGGTGACAACAGATATGACTCTGGAUGGAAT	2377
QY	2385	TAATATAAAGAAGATATGTTCCCAAGTTCCAACAAGCCCAATAACAGATGTGCATTTCTT	2444
Db	2378	CACCTCCCCAGCTGAACCTTTTCCACCTGGAGTCCTTGGGATACCGGACGTGATCTCTT	2437
QY	2445	TTATAAGTCTTCTACAGCAACAACATCTTGTATTAAATGSCCGATCAACTGCTGTGAAAAT	2504
Db	2438	TTATAGGTCCAATGATGTGACCCAGTCTCTGCAGTCTTGGGAGATCAACCACCATCCGGT	2497
QY	2505	GAGTGTAACTCCTACTAAATCTGGAGCAGGAGTGATTTTCACTCCCCAGCAAGTGCCACGC	2564
Db	2498	CAGGTGCAGTCCACAGAAAACCTGTCCCTGGAAGTTTGTCTGCCAGGAACGTGCTCAGA	2557
QY	2565	AGGTACCTGTGATGGGTGTACGTTCTATTTCTCTGTGGGAGAGTGTGAAGCTTGCCCTCT	2624
Db	2558	TGGGACCTGTGATGGCTGCAACTTCCACTTCTCTGTGGGAGAGCGCGGTGCTTGCCCGCT	2617
QY	2625	GTGTACGGAGCATGACTTCCATGAGATTTGAGGGAGCCTGCAAGAGAGGATTTTCAGGAAC	2684
Db	2618	CTGCTCAGTGGCTGACTACCGTGCTATCGTCAGCAGCTGTGTGGTGGGATCCAGAAGAC	2677
QY	2685	CTTGATGTGTGAATGAACCTAAATGGTGCATTAAGGAATTTCTTTGCCTGAGAAAAA	2744
Db	2678	TACTTACGTGTGGCGAGAACCACCAAGCTATGCTCTGGTGGCATTTCTCTGCCTGAGCAG	2737
QY	2745	GTTGGCAACCTGTGAAACCGGTTGACTTTTGGCTGAAGGTGGAGCGCGGTGTGGGAGCTTT	2804
Db	2738	AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGSCATCTCTGCAGGCCCTG	2797
QY	2805	TACTGCCGTTTGTGGTGGCTCTGACCTGCTACTTCTGAAAAAAGAAATCAAAAAGAAAA	2864
Db	2798	TACTGCCATCCTGCTCACCGTCTTGACCTGCTACTTTTGGAAAAAGAAATCAAAAACCTAG	2857
QY	2865	GAAGACCAATTTTGAATCTGTTCA 2887	
Db	2858	GTACAAGTACTCCAAGCTGGTGA 2880	
RESULT 6			
AX463904			
LOCUS			
DEFINITION			
AX463904			
ACCESSION			
AX463904.1			
VERSION			
AX463904.1			
KEYWORDS			
SOURCE			
human.			
ORGANISM			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE			
1			
AUTHORS			
Baker,K.P., Beresini,M., Deforge,L., Desnoyers,L., Filvaroff,E.,			
Gao,W.Q., Gerritsen,M.E., Goddard,A., Godowski,P.J., Gurney,A.L.,			
Sherwood,S., Smith,V., Stewart,T.A., Tumas,D., Watanabe,C.K.,			
Wood,W.L. and Zhang,Z.			
TITLE			
Secreted and transmembrane polypeptides and nucleic acids encoding			
same			
JOURNAL			
Patent: WO 0140466-A 37 07-JUN-2001;			
Genentech Inc. (US)			
FEATURES			
source			
1..3501			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
BASE COUNT			
921 a 979 c 823 g 776 t			
OTHERS			
2 others			
Query Match			
26.6%; Score 901; DB 6; Length 3501;			
Best Local Similarity 60.0%; Pred. No. 3.8e-200;			
Matches 1634; Conservative 0; Mismatches 1062; Indels 27; Gaps 7;			
QY	171	GCTTCTCTCTTGGCAGGAGAAAGATTATACCTTTGAATATACGGAATGTGATAGCAGTGG	230



Db 267 GCITCATGCCTGCAAAAGAGTCTGAGTACCACCTATGAGTACACGGCGGTGTGACAGCACGGG 326

QY 231 CTCAGGTGGAGAGTTGCCAATCCAAAAATCTGCAGTGGACTGCTCTGGCCTGCGCTGAGCC 290

Db 327 TTCAGGTGGAGGTGCGCGTGCCGCATACCCCGGCCTGTGCACACGCTGTCTGACCC 386

QY 291 AGTGAGAGGCAAAAGAATGCACTTTCTCCTGTGCTTCTGGAGAGTATCTAGAAAATGAAGAA 350

Db 387 CGTCAAGGGCACCGAGTGCCTCTCTCCTGCAACGCCGGGAGTTTCTGGATATGAAGGA 446

QY 351 CCAGGTATGCAGTAAGTGTGGTGAAGGCACCTATTCTCTGGGCAGTGGCATCAAAATTGA 410

Db 447 CCAGTCATGTAAGCCATGCGCTGAGGCGCGCTACTCCCTCGGCACAGGCAATTCGGTTGA 506

QY 411 TGAATGGGATGAATTGCCGCGAGGATTTTCTAAACATCGCAACATTCATGGACACTGTGGT 470

Db 507 TGACTGGGATGAGCTGCCCATGGCTTTGCCAGCCTCTCAGCCAAACATGGAGCTGGATGA 566

QY 471 GSGCCCTTCTGACAGCAGGCCAGACGGCTGTAACAACTCTTCTTGGATCCCTCGTGGAAA 530

Db 567 CAGTGTCTGAGTCCA---CCGGGAAGTGTACTTTCGTCCAAGTGGGTTCGCCGGGCGGA 623

QY 531 CTACATAGAATCTAATCGTGTGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAA 590

Db 624 CTACATCGCCTCCAACACGGAGCAATGCACGCCACACTGATGTACGCCGTCAACCTGAA 683

QY 591 GAAGTCAAGGCTATGTCTTCTTTGAGTACCAGTGTCTTTGATCTATGCTGTGCACCTTAA 650

Db 684 GCAATCTGGCACCCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 743

QY 651 CTTTATTCAAAATGATCAGTGCCAGGAGATGGACACCCACACTGACAAGTGGGTAAAACT 710

Db 744 TTTGCTCAGAATGACCAAGTGCCAGCCCCAATGCAGATGACTC---CAGGTGGATGRAGAC 800

QY 711 TACAGACAAATGGAGAATGSGGCTCTCATTTCTGTAATGCTGAATCAGGCACAAACATACT 770

Db 801 CACAGAGAA--AGGATGGAAATCCACAGTGTGGAGCTAAATCGAGGCAATAATGTCT 857

QY 771 CTACTGGAGAACTACAGGCATCCTTATGGGTTCTAAGCGGTCAAGCCTGTGCTGTAA 830

Db 858 CTATTGGAGAACCCACAGCCTTCTCAGTATGGACCAAGTACCCAAAGCCTGTGTGTGAG 917

QY 831 AAATATCACAAATTGAAGGGTGGCGGTACACATCAGAAATGTTTCTTGCAAGCCAGGCAC 890

Db 918 AAACATTGCCATAACAGGGTGGCCTACACTTCAGAATGCTTCCCCTGCAAAACCTTGGCAC 977

QY 891 ATTCAGCAACAAACCAGGTTCAATCAACTGCCAGGTGTGTCCCAGAAACACCTATTCTGA 950

Db 978 GTATGCAGACAAGCAGGGCTCCTTTCTGCAAACTTTGCCAGCCAACCTCTTTATTCAAA 1037

QY 951 GAAAGGAGCCAAAGAAATGTATAAGGTGTAAA---GAGCACTCTCAATTTTCAGGATCCAG 1007

Db 1038 TAAAGGAGAAACTTCTTGCCACCAGTGTGACCCCTGACAAATACTCAGAGAAAGGATCTTC 1097

QY 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATCTCCATG 1067

Db 1098 TTCCTGTAACGTGCGCCCGCTTGACACAGACAAAGATTATTTCTACACACACCGGCCTG 1157

QY 1068 TGATGAAGAAGGAAAGACACAGATAATGTACAAAGTGGATAGAGCCCCAAAATCTGCCGGGA 1127

Db 1158 CGATGCCAACGGAGAGACAACTCATGTACAAAATGGCCAAAGCCGAAAATCTGTAGCGGA 1217

QY 1128 GGATCTCACAGATGCTATTAGATTTGCCCTTCTGGAGAGAAAGAGGATTGTCCGCTTG 1187

Db 1218 GGACCTTGAGGGGCGAGTGAAGCTGCCTGCCTCTGCTGTGAAGACCCCACTGCCCAACCTG 1277

QY 1188 CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCCCTGGAACATT 1247

Db 1278 CAACCCAGGCTTCTTCAAAACCAACAACAGCACCTGCCAGCCCTGCCCATATGGTTCCCTA 1337

QY 1248 TTCAGATGGAACCAAGAAATGTAGACCATTGTCCAGCAGGAACGGAGCCTGCACCTTGGCTT 1307

Db 1338 CTCCAATGG---CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGTCTGTGGGAT 1394

QY 1308 TGAATATAAATGGTGGAAATGTCTTCTCCTGGCAACATGAAAACCTTCTCTCAATGTTGG 1367

Db 1395 TGAATACAAATGGTGGAAACACGCTGCCACAAACATGGAACGACCGTTCTCAGTGGGAT 1454

QY 1368 GAATTCAAAAGTCCGATGGAATGAATGGTTGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427

Db 1455 CAACTTCGAGTACAAGGGCATGACAGGCTGGGAGGTGGTGGTGTGATCACATTTACACAGC 1514

QY 1428 GGTGGAGGTTCTGACAAATGATPACCTGATCTTAAACTTGCATATCCCAGGATTTAAACC 1487

Db 1515 TGCTGGAGCCTCAGACAATGACTTTCATGATTTCTCACTCTGTTGTGCCAGGATTTAGACC 1574

QY 1488 ACCAACATCTATGACTGGAGCCAC--GGGTTCTGAACTAGGAAGAATAACATTTGTCTT 1544

Db 1575 TCCGCAGTGGTGTATGGCAGACACAGAGAATAAAGAGGTGGCCAGAATCACATTTGTCTT 1634

QY 1545 TGAGACCCCTCTGTTTCAGCTGACTGTGTTTGTACTTTCATGTTGGATATTAATAGAAAAAG 1604

Db 1635 TGAGACCCCTCTGTTCTGTGAACCTGTGAGCTCTACTTTCATGTTGGTGTGAATTTCTAGGAC 1694

QY 1605 TACAAATGTGGTAGAATCGTGGGTGGAACCAAGAAAAAGAAACAAAGCTTACACCCATATCAT 1664

Db 1695 CAACACTCCTGTGGAGACGTGGAAGGTTCCAAAGGCAACAGTCTATACCTACATCAT 1754

QY 1665 CTTCAAGAAATGCAACTTTTACATTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGA 1724

Db 1755 TGAGGAGAACACTACCACGAGCTTCACCTGGGCCTTCCAGAGACCACTTTTCATGAGGC 1814

QY 1725 TAATAGACGGTTTCATCAATGACATGGTGAAGATTTATTTCTATCACAGCCACTAATGCAGT 1784

Db 1815 AAGCAGGAAGTACACCAATGACGTGCGCAAGATCTACTCCATCAATGTACCAATGTTAT 1874

QY 1785 TGATGGGTGGCGTCCCTCATGCCGTGCCCTGTGCCCTCGTTCGTAACAGTCGGGTTTCATC 1844

Db 1875 GAATGGCGTGGCCTCCTACTGCCGTCCCTGTGCCCTAGAAGCCTCTGATGTGGGCTCCTC 1934

QY 1845 GTGTGTCCCTGCCCTCCAGGCCACTACATTTGAGAGAAAGAAACCAACCAAGTGCAGGAAATG 1904

Db 1935 CTGCACCTCTTGTCTCTGCTGGTTACTATATTGACCGAGATTTCAGGAACCTGCCACTCCCTG 1994

QY 1905 TCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAGAGGGCTTGTATTTCCATG 1964

Db 1995 CCCCCCTAACACAAATCTGAAAAGCCCAACAGCCTTATGGTGTCCAGGCCTGTGTGCCCTG 2054

QY 1965 CGGCCTGGGAGTAAAAACAATCAGGACCAATTCGGTTTGTCTATAGTGAATGCTTTTCTTA 2024

Db 2055 TGSTCCAGGGACCAAGAACAAAGATCCACTCTCTGTGCTACAATGATTGCACCTTCTC 2114

QY 2025 CCATGAAAAAGAAAATCAGATTTTGCACTATGACTTTTAGCAACCTCAGCAGTGTGGGCTC 2084

Db 2115 ACGCAACACTCCAACCCAGGACTTTCAACTACAACCTTCTCCGCTTGGCAACACCGCTCAC 2174

QY 2085 ATTAATGAATGGCCCCAGCTTCACTTCCAAAGGAACAAAATACTTCCATTTCTTCAATAT 2144

Db 2175 TCTTGTGGAGGGCCAAGCTTCACTTCCAAAGGTTGAATACTTCCATCACTTTACCCCT 2234

QY 2145 CAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATAATAACAGACTT 2204

Db 2235 CAGTCTCTGTGGAAACCAGGGTAGGAAAATGTCTGTGTGACCGGACAAATGTCACTGACCT 2294

QY 2205 TACAGTAAAAGAAATAGTGGCAGGTTCAGATGATTACACAAAATTTGGTAGGGGCATTTGT 2264

Db 2295 CCGGATTCCTGAGGGTGAGTCAGGG-----TTCCTCAAAATCTATCACAGCCTACGT 2345

QY 2265 ATGCCAGTCAACAAATTAATCCTTCTGAAAAGTAAGGGTTTCGAGCAGCCTTATCATCACA 2324

Db 2346 CTGCCAGGCAGTCATCATCCCCCAGAGGTGACAGGCTACAAGGCGGGGTTTCCTCACA 2405

QY 2325 ATCCATCATTTCTGGCAGATACATTTCATAGGAGTCACAGTTCAGAACCAACATTTGAAAAATAT 2384

Db 2406 GCCTGTACGCCCTTGCTGTATCCACTTATTGGGGTGACAACAGATATGACTCTGGATGGAAT 2465











Db	1983	ACGCAACACTCCAACAGGACTTTTCAACTACAACCTCTCCGCTTTGGCAAAACACCGTCAC	2042
QY	2085	ATTAATGAATGSCCCAGCTTCACCTCCAAAGGAACAAAATACTTCCATTCTTCAATAT	2144
Db	2043	TCTTGCTGGAGGCCAAGCTTCACCTTCCAAAGGGTTGAAATACTTCCATCACTTACCCT	2102
QY	2145	CAGTTTATGTGGGCATGAGGGGAAGAAGATGSGCTCTCTGTACCAACAATATAACAGACTT	2204
Db	2103	CAGTCTCTGTGGAACACAGGGTAGGAAAATGTCGTGTGCACCGACAATGTCACCTGACCT	2162
QY	2205	TACAGTAAAAGAAATAGTGGCAGGGTCAGATGATTACACAAAATTTGGTAGGGGCATTTGT	2264
Db	2163	CCGGATTCTGAGGGTGAGTCAGGGTT-----CTCCAAATCTATCACAGCCTACGT	2213
QY	2265	ATGCCAGTCAACAATTATTCTCTGAAAGTAAGGGTTTCCGAGCAGCGCTTATCATCACA	2324
Db	2214	CTGCCAGGCAGTCATATCCCCCAGAGGTGACAGGCTACAAGGCCGGGGTTCTCTCACA	2273
QY	2325	ATCCATCATTTCTGGCAGATACATTTCATAGGAGTCACAGTTGAAACCACATTTGAAAAATAT	2384
Db	2274	GCCTGTGAGCCTTGCTGATCGACTTATTGGGGTGACAACAGATATGACTCTGGATGGAAT	2333
QY	2385	TAATATAAAGAAGATATGTTCCAGTTTCCAAACAAGCCAAAATACCAGATGTGCATTTCTT	2444
Db	2334	CACCTCCCAGCTGAACTTTTCACCTGGAGTCCTTGGGAATACCCGGACGTGATCTTCTT	2393
QY	2445	TTATAAGTCTTCTACAGCAACAACATCTTGTATTAAATGGCCGATCAACTGCTGTGAAAAT	2504
Db	2394	TTATAGGTCCAATGATGTGACCCAGTCCTGCAGTCTCTGGGAGATCAACCACCATCCGCGT	2453
QY	2505	GAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCCCAAGTGTCCCCAGC	2564
Db	2454	CAGGTGCAGTCCACAGAAAACGTCCCTGGAAAGTTTGCTGTGCCAGGAACGTGCTCAGA	2513
QY	2565	AGGTACCTGTGATGGGTGACGTTCTATTTCCTGTGGGAGAGTGTCTGAAGCTTGCCCTCT	2624
Db	2514	TGGGACCTGTGATGGCTGCAACTTCCACTTCCTGTGGGAGAGCGCGGTGCTTGCCCGCT	2573
QY	2625	GTGTACGGAGCATGACTTCCATGAGATTTGAGGGAGCCTGCAAGAGAGAGGATTCAGGAAC	2684
Db	2574	CTGCTCAGTGGCTGACTACCGTGTCTATCGTCAGCAGCTGTGTGGCTGGATCCAGAAGAC	2633
QY	2685	CTTGATGTGTGGAATGAACCTAAATGGTGCAATTAAGGAATTTCTTTGCCCTGAGAAAAA	2744
Db	2634	TACTTACGTGTGGCGAGAACCCAAAGCTATGCTCTGTGGGCATTTCTCTGCCTGAGCAGAG	2693
QY	2745	GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAAAGTG	2784
Db	2694	AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTG	2733
RESULT	9		
AK057647			
LOCUS	AK057647	2895 bp	mRNA
DEFINITION	Homo sapiens cDNA FLJ33085 fis, clone TRACH2000420, weakly similar to MAJOR SURFACE-LABELLED TROPHOZOITE ANTIGEN PRECURSOR.	linear	PRI 01-AUG-2002
ACCESSION	AK057647		
VERSION	AK057647.1	GI:16553406	
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens trachea cDNA to mRNA, clone_lib:TRACH2 clone:TRACH2000420.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 Suzuki,O., Sasaki,N., Aotsuka,S., Shoji,T., Ichihara,T., Shiohata,N., Matsumoto,K., Hirano,M., Sano,S., Nomura,R., Yoshikawa,Y., Matsumura,Y., Moriya,S., Chiba,E., Momiyama,H., Onogawa,S., Kaeriyama,S., Satoh,N., Matsunawa,H., Takahashi,E., Kataoka,R., Kuga,N., Kuroda,A., Satoh,I., Kamata,K., Takami,S., Terashima,Y., Watanabe,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,		
	Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.		
	NEDO human cDNA sequencing project		
	Unpublished		
	2 (bases 1 to 2895)		
	Isogai,T., Otsuki,T. and Sugiyama,T.		
	Direct Submission		
	Submitted (24-Oct-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)		
	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.		
FEATURES	Location/Qualifiers		
Source	1. .2895		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="TRACH2000420"		
	/tissue_type="trachea"		
	/clone_lib="TRACH2"		
	/note="cloning vector: pME18SFL3"		
BASE COUNT	692 a 825 c 714 g 664 t		
ORIGIN			
	Query Match 20.9%; Score 709.6; DB 9; Length 2895;		
	Best Local Similarity 57.0%; Pred. No. 2.7e-155;		
	Matches 1550; Conservative 0; Mismatches 999; Indels 171; Gaps 7;		
QY	171	GCTTCCTCCTTGCCAGGAGAAAGATTATCACCTTGAATATACGGAATGTGATAGCAGTGG	230
Db	214	GCTTCATGCCTGCAAAAGAGTCTGAGTACCACATATGAGTACACGGCGTGTGACAGCACGGG	273
QY	231	CTCCAGGTGGAGAGTTGCCATTCCAAATCTGCAGTGGACTGCTCTGGCCTGCCTGACCC	290
Db	274	TTCCAGGTGGAGGGTCGCGGTGCCGATACCCCGGCGCTGTGCACCAGCCTGCCTGACCC	333
QY	291	AGTGAGAGGCCAAAGATGCACATTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAA	350
Db	334	CGTCAAGGGCACCGAGTGTCTCTCCTGCAACGCCGGGGAGTTCTGGATATGAAGGA	393
QY	351	CCAGGTATGCAGTAAGTGTGGTGAAGGCACCTATTCCCTGGGCAGTGGCATCAAAATTGA	410
Db	394	CCAGTCATGTAAGCCATGCGCTGAGGGCGGCTACTCCCTCGGCACACAGGCATTCGGTTGA	453
QY	411	TGAATGGGATGAATTGCCGGCAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGT	470
Db	454	TGAGTGGGATGAGCTGCCCATGGCTTTGCCAGCCTCTCAGCCAACATGGAGCTGGATGA	513
QY	471	GGGCCCTTCTGACAGCAGGCCAGCGGCTGTAACAACACTCTTCTTGGATCCCTCGTGGAAA	530
Db	514	CAGTGTCTGTGAGTCCA---CCGGGAACGTGACTTCGTCCAAGTGGGTCCCCCGGGCGGA	570
QY	531	CTACATAGAAATCTAATCGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAA	590
Db	571	CTACATCGCCTCCAACACACGGACGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA	630
QY	591	GAAGTCAGGCTATGTCTTCTTTGAGTACCAGTATGTGCACAACAACATCTTCTTTGAGTT	650
Db	631	GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT	690
QY	651	CTTTATTCAAAATGATCAGTGCCAGGAGATGGACACCACCACTGACAAAGTGGGTAAACT	710
Db	691	TTTCGTTCAGAATGACCACTGCCAGCCCAATGCAGATGACTC---CAGGTGGATGAAGAC	747
QY	711	TACAGACAATGGGAATGGGCGCTCTCATTTCTGTAAATGCTGAAATCAGGCACAAACATACT	770



Db 748 CACAGAGAA---AGGATGGGAATTCACAGAGTGTGGAGCTAAATCGAGGCAATAATGTCTCT 804

Qy 771 CTACTGGAGAACTACAGGCATCCTTATGGGTTCCTAAGGGGTCAAGCCTGTGCTGGTAAA 830

Db 805 CTATTGGAGAACACAGACCTTCTCAGTATGGACCACAAAGTACCAAAGCCTGTGCTGGTGAG 864

Qy 831 AAATATCACAATTAAGGGGTGGCGTACACATCAGAATGTTTTCTCTTGCAAAGCCAGGCAC 890

Db 865 AAACATTGCCATAACAGAG----- 883

Qy 891 ATTACAGCAACAAACCAGGTTCAATTCAACTGCCAGGTGTGTCACAGAAACACCTATTCTGA 950

Db 884 ----- 883

Qy 951 GAAAGGAGCCAAAGAATGTATAAGSTGTAAAGACGACTCTCAATTTTCAGGATCCAGTGA 1010

Db 884 -----AAGGATCTTCTTC 897

Qy 1011 GTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATACTCCATGTGA 1070

Db 898 CTGTAAGCTGGCCCCAGCTTGCACAGACAAAGATTATTTCTACACACACACAGGCCCTGCCGA 957

Qy 1071 TGAAGAACGAAGACACAGATATGTACAAAGTGGATAGAGCCCAAAAATCTGCCGGGAGGA 1130

Db 958 TGCCAACGGAGAGACACAACCTCATGTACAAATGGGCCAAGCCGAAANAATCTGTAGCGAGGA 1017

Qy 1131 TCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAGAAGGATGTCCCGCCTTGCAA 1190

Db 1018 CCTTGAGGGGGCAGTGAAGCTGCCTGCCTCTGGTGTGAAGACCCCACTGCCACCCCTGCAA 1077

Qy 1191 CCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTCTGGAACATTTTC 1250

Db 1078 CCCAGGCTTCTTCAAAACCAACAACAGCACCTGCCAGCCCTGCCCATATGTTTCTCTACTC 1137

Qy 1251 AGATGGAACCAAGAANTGTAGACCATGTCCAGCAGGAACGAGGCCTGCACCTTGGCTTTGA 1310

Db 1138 CAATGG---CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAAACCTGCTGTGGGATTGA 1194

Qy 1311 ATATAAATGGTGAATGTCTTCCTGGCAACATGAAAACCTTCCTGCTTCAATGTTGGGAA 1370

Db 1195 ATACAAATGGTGAACACCGCTGCCACAAACATGGAACGACCGTTCCTCAGTGGGATCAA 1254

Qy 1371 TTCAAAGTCGGATGGAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGC 1430

Db 1255 CTTTCGACTACAAGGCATGACAGCTGGGAGGTGGCTGGTGATCACATTTACACAGCTGC 1314

Qy 1431 TGGAGGTTCTGACAAATGATTACCTGATCTTAAACTTGCATATCCCAGGATTTAAACCACC 1490

Db 1315 TGGAGCCTCAGACAAATGACTTCATGATTTCTCACTCTGGTGTGCCAGGATTTAGACCTCC 1374

Qy 1491 AACATCTATGACTGGAGCCAC---GGGTTCTGAACCTAGGAAGAATAACATTTGTCTTTGA 1547

Db 1375 GCAGTCGGTGATGGCAGACACAGAGAATAAAGAGGTGGCCAGAATCACATTTGTCTTTGA 1434

Qy 1548 GACCTCTGTTCAGCTGACTGTGTTTTGTACTTTCATGFTGGATATTAATAGAAAAAGTAC 1607

Db 1435 GACCTCTGTCTGTGAACCTGTGAGCTCTACTTTCATGTGGGTGTGAATTTCTAGGACCAA 1494

Qy 1608 AAATCTGGTAGAATCGTGGGTGGAAACCAAGAAACAAAGCTTACACCCATATCATCTTT 1667

Db 1495 CACTCCTGTGGAGACGTGGAAAGGTTCCAAAGGGCAACAGTCCCTATACCTACATCATGA 1554

Qy 1668 CAAGAATGCAACTTTTACATTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGATAA 1727

Db 1555 GGAGAACACTACCACGAGCTTCACCTGGGCTTCCAGAGGACCACTTTTTCATGAGGCAAG 1614

Qy 1728 TAGACGGTTTCATCAATGACATGGTGAAGATTTTATTTCTATCACAGCCCACTAATGCACTGA 1787

Db 1615 CAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAGTGTCAACCAATGTTATGAA 1674

Qy 1788 TGGGGTGGCGTCCCTCATGCGCTGCCCTGTGCCCTCGGTTCTGAACAGTTCGGGTTTCATCGTG 1847

Db 1675 TGGCGTGGCCTCCTACTGCGCTCCCTGTGCCCTAGAGCCCTCTGATGTGGGCTCCTCCTG 1734

Qy 1848 TGTCCCTGCCCCTCCAGGCCCACTACATTGAGAAAGAAACCAACCAGTGCAAGGAATGTCTCC 1907

Db 1735 CACCTCTTGTCTCTGCTGGTTACTATATTTGACCGAGATTTCAGGAACCTGCCACTCTCTCC 1794

Qy 1908 ACCTGACACACCTACCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTCATCCGCG 1967

Db 1795 CCCTAACACAAATTCGAAAGCCCAACAGCCCTTATGGTGTCCAGGCCCTGTGTGCCCTGTGG 1854

Qy 1968 GCTGGGAGTAAAAACAATCAGGACCAATTCGGTTTGTATAGTACTGCTTTTCTPACCA 2027

Db 1855 TCCAGGGACCAAGAACAACAGATCCACTCTCTGTGCTACAAATGATTGCACCTTCTCACG 1914

Qy 2028 TGAANAAGAAATCAGATTTTGCACATATGACTTTAGCAACCTCAGCAGTGTGGGCTCATTT 2087

Db 1915 CAACACTCCAACCCAGGACTTTCAACTACAACCTCTCCGCTTTGGCAACACCCGTCACCTCT 1974

Qy 2088 AATGAATGGCCCCAGCTTCACCTCCAAAGGAACAANAATACTTCCATTTCTTCAATATCAG 2147

Db 1975 TGTGGAGGGCCCAAGCTTCACCTTCCAAAGGGTTGAATACTTCCATCACCTTTACCCTCAG 2034

Qy 2148 TTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTGTACCAACAATAAACAAGACTTTTAC 2207

Db 2035 TCTCTGTGGAACCCAGGGTAGGAAAATGTCTGTGTGACCCGACAATGTCTCAGTACCTCCG 2094

Qy 2208 AGTAAAGAAATAGTGGCAGGGTCAAGATGATTACACAAATTTGGTAGGGCAATTTGTATG 2267

Db 2095 GATTCCCTGAGGGTGAGTCAAGG-----TTCCTCCAAATCTATCACAGCCTACGTTCTG 2145

Qy 2268 CCAGTCAACAATTTATTCCTTCTGAAAGTAAGGGTTTCCGAGCAGCCCTTATCATCACAAATC 2327

Db 2146 CCAGGCAGTCATCATCCCCCAGAGGTGACAGGCTACAAAGCCGGGTTTCTCTCACAGCC 2205

Qy 2328 CATCATTTCTGGCAGATACATTCATAGGAGTCACAGTTGAAACCACATTTGAAAATAATTA 2387

Db 2206 TGTACAGCCTTGTGATCGACTTATTGGGGTGACAAACAGATATGACTCTGGATGGAATCAC 2265

Qy 2388 TATAAAGAAAGATATGTTCCCAAGTTTCCAAACAAGCCAAATACCAGATGTGCATTTCTTTA 2447

Db 2265 CTCCCCAGCTGAACCTTTCCACCTGGAGTCCCTTGGGAATACCGGACGTGATCTTCTTTA 2325

Qy 2448 TAAGTCTTCTACAGCAACAACATCTTGTATTAAATGCCCCGATCAACTGCTGTGAAAATGAG 2507

Db 2326 TAGGTCCAATGATGTGACCCAGTCTCTGCAGTTCTGCGGAGATCAACCCACCATCCGCGTCAG 2385

Qy 2508 GTGTAATCCTACTAAATCTGGAGCAGGAGTGTATTCAGTCCCCAGCAAGTGCCCAGCAGG 2567

Db 2385 GTGCAGTCCACAGAAAACCTGTCCCTGGAAAGTTTGTGCTGCCAGGAACGTGCTCAGATGG 2445

Qy 2568 TACCTGTGATGGGTGACGTTCTATTTCTCTGTGGGAGAGTGTGAAAGCTTGCCCTCTGTG 2627

Db 2446 GACCTGTGATGGCTGCAACTTCCACTTCCCTGTGGGAGAGCGCGGCTGCTTGCCCGCTCTG 2505

Qy 2628 TACGGAGCATGACTTCCATGAGATTTGAGGGAGCCTGCAAGAGAGGATTTTCAGGAACCTTT 2687

Db 2506 CTCAGTGGCTGACTACCATGCTATCGTCAAGCAGCTGTGTGGTGGATCCAGAAAGACTAC 2565

Qy 2688 GTATGTGTGGAATGAACCTAAATGGTGCATTTAAAGGAATTTCTTTGCCCTGAGAAAAAGTT 2747

Db 2565 TTACGTGTGGCGAGAACCCCAAGCTATGCTCTGCTGGTGCATTTCTCTGCTGAGCAGAGAGT 2625

Qy 2748 GGCAACCTGTGAAACGTTGACTTTTGGCTGAAGTGGGAGCCGGTGTGGGAGCTTTTAC 2807

Db 2626 CACCATCTGCAAAACCATAGATTTCTGGCTGAAGTGGGCATCTCTGTCAGGSCACCTGTAC 2685

Qy 2808 TGCCGTTTTGCTGGTGCCTCTGACCTGTACTTTCTGGAAAAAAGAANTCAAAAAGAAAAAGAA 2867

Db 2685 TGCCATCTCTGCTCACCGTCTTGACCTGTACTTTTGGAAAAAAGAATCAAAAAGACTAGAGTA 2745

Qy 2868 GACCATTTTGAATCTGTCA 2887

Db 2746 CAAGTACTCCAAGCTGGTGA 2765







Db	630	CCCTAGAAGCCTCTGATGTGGGCTCCTCCTGTCACCTCTTGTCTCCTGCTGGTTACTATATATG	689
Qy	1877	AGAAAGAAACCAACCAAGTGCAGGAATGTCCACCTGACACCTACCTGTCCATACATCAAGG	1936
Db	690	ACCGAGATTGAGAAACCTGCCACTCCTGCCCCCTAACACAATTCTGAAAGCCCAACAGC	749
Qy	1937	TCTATGGCAAAAGAGGCTTGTAFTCCATGCGGGCCTGGGAGTAAAAACAATCAGGACCAT	1996
Db	750	CTTATGGTGTCAGGCCTGTGTGCCCTGTGTCTCCAGGACCAAGAACAACAAGATCCACT	809
Qy	1997	CGGTTTGCTATAGTGACTGCTTTTCTTACCATGAAAAAGAAAAATCAGATTTTGCACATG	2056
Db	810	CTCTGTGCTACAACGATTGCACCTTCTCACGCAACACTCCGACCAGGACTTTCACACTACA	869
Qy	2057	ACTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGSCCCCCAGCTTCACCTCCAAAG	2116
Db	870	ACTTCTCCGCTTTGGCAAAACACTGTCACTCTTGCTGGAGGCCCAAGCTTCACCTTCCAAAG	929
Qy	2117	GAACAAAATACTTCCATTTCTCAATATATCAGTTTATGTGGCATGAGGGGAAGAAGATGG	2176
Db	930	GGCTGAANAATCTTCCATCACTTTACCCTCAGTCTCTGTGGAAACCAGGGTAGGAAAATGT	989
Qy	2177	CTCTCTGTACCAACAATATAACAGACTTTTACAGTAAAGAAATAGTGGCAGGGTCAGATG	2236
Db	990	CTGTGTGCACCGACAATGTCACTGACCTCCGGATTCTGTGAGGGTGAGTCAGGG-----	1042
Qy	2237	ATTACACAAAATTTGGTAGGGCATTTGTATGCCAGTCAACAATATTCTTCTGAAAGTA	2296
Db	1043	--TTCTCCAAATCTATCACAGCCTACGTCTGCCAGGCAGTCATCATCCCCCAGAGGTGA	1100
Qy	2297	AGGGTTTCCGAGCAGCCTTATCATCACAATCCATCATTTCTGGCAGATACATTTCATAGSAG	2356
Db	1101	CAGGCTACAAGGCCGGGTTTCTCTCACAGCCTGTGCAGCCTTGCTGATCGACTTATTGGGG	1160
Qy	2357	TCACAGTTGAAACCAACATTTGAAAAATATTAATATAAAAGAAAGATATGTTCCCAAGTCCAA	2416
Db	1161	TGACACAGATATGACTCTGGATGGAATCACCTCCCCAGCTGAACTTTTCCACCTGGAGT	1220
Qy	2417	CAAGCCAAATACCAGATGTGCATTTCTTTTATAAGTCTTCTACAGCAACAACATCTTGTA	2476
Db	1221	CCTTGGGAATACCGGACGTGATCTTCTTTTATAGTCCAAATGATGTGACCCAGTCCTGCA	1280
Qy	2477	TTAATGGCCGATCAACTGCTGTGAAAAATGAGGTGTAATCCTACTAAATCTGGAGCAGGAG	2536
Db	1281	GTTCTGGGAGATCAACCAACCATCCGCGTCAGGTGCAGTCCACAGAAAACTGTCCCTGGAA	1340
Qy	2537	TGATTTCAGTCCCAGCAAGTGCCCAGCAGGTACCTGTGTGATGGGTGTACGTTCTATTTC	2596
Db	1341	GTTTGTGCTGCCAGGAACGTGCTCAGATGGGACCTGTGTGCTGCTGCAACTTCCACITCC	1400
Qy	2597	TGTGGGAGAGTGCTGAAGCTTGGCCCTCTGTGTACGGAGCATGACTTCCATGAGATTGAGG	2656
Db	1401	TGTGGGAGAGCGCGGTGCTTGGCCGCTCTGCTCAGTGGCTGACTACCATGCTATCGTCA	1460
Qy	2657	GAGCCTGCAAGAGAGGATTTTCAGAAACCTTGTATGTGTGGAATGAACCTAAATGGTGCA	2716
Db	1461	GCAGCTGTGTGGCTGGGATCCAGAAGACTACTTACGTGTGGCGAGAACCCCAAGCTATGCT	1520
Qy	2717	TTAAGGAAATTTCTTGGCCTGAGAAAAAGTTGGCAACCTGTGAAACGGTTGACTTTTGGC	2776
Db	1521	CTGGTGGCATTTCTCTGCCTGAGCAGAGAGTCACCATCTGCAAAAACCATAGATTTTCTGGC	1580
Qy	2777	TGAAGTGGGAGCCGGTGTGGGAGCTTTTACTGCCGTTTTTGTGGTGGCTCTGACCTGCT	2836
Db	1581	TGAAAGTGGGCATCTCTGCAGGCACCTGTACTGCCATCCTCTGCACCGTCTTGACCTGCT	1640
Qy	2837	ACTTCTGAAAAAAGAAATCAAAAGAAAAAGAACCATTTTGAATCTGTTCA	2887
Db	1641	ACTTTTGGAAAAAGAAATCAAAAACACTAGAGTACAAGTACTCCACAGCTGGTGA	1691

RESULT 12  
AC002081/c

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
  
COMMENT

AC002081 136150 bp DNA linear PRI 21-DEC-1999  
Homo sapiens BAC clone CTA-331C24 from 7q21, complete sequence.  
AC002081  
AC002081.1 GI:2078453  
HTG.  
Homo sapiens.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 136150)  
Du,Z. and Maggi,L.  
The sequence of Homo sapiens BAC clone CTA-331C24  
Unpublished  
2 (bases 1 to 136150)  
Waterston,R.  
Direct Submission  
Submitted (12-MAY-1997) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
3 (bases 1 to 136150)  
Waterston,R.  
Direct Submission  
Submitted (24-SEP-1998) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
4 (bases 1 to 136150)  
Waterston,R.  
Direct Submission  
Submitted (21-DEC-1999) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
Center project name: H\_RG331C24  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GTB/CHR7> or <mailto:egreen@nhgri.nih.gov>

SOURCE INFORMATION:

Clone CTA-331C24 is from a release of the human BAC library CITB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).  
VECTOR: pBelobAC11  
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is CTA-369H2, 200 bp overlap.  
Actual start of this clone is at base position 1 of CTA-331C24;  
actual end is at 136150 of CTA-331C24.

FEATURES

source

1..136150  
/organism="Homo sapiens"





Db 63236 GGCCATGCCGCTGAAAAGGGAAGGAGATTGAAACATTGATTCGCTTATCACATGGTCAA 63177

QY 3093 GTACCTTGCCAAATAAAGGAAAGCAAATGATTTGGGTCTCAAACTGAAGATGAAGCTCAAC 3152

Db 63176 GTACCTTGCCAAATAAAGGAAAGCAAATGATTTGGGTCTCAAACTGAAGATGAAGCTCAAC 63117

QY 3153 TCAGGAAGAGATTTATCTGTATATACACATAACTGAAACCAAAAGTTTAAAGCCCACCAATG 3212

Db 63116 TCAGGAAGAGATTTATCTGTATATACACATAACTGAAACCAAAAGTTTAAAGCCCACCAATG 63057

QY 3213 CACTGCTGATGCATGCCATATAATTAATGGGTAACITTTATCTTTATGATGTCTACATA 3272

Db 63056 CACTGCTGATGCATGCCATATAATTAATGGGTAACITTTATCTTTATGATGTCTACATA 62997

QY 3273 ACAAGTGTGATTTGGAAGGCACATGTGAGCATATGCATATATGATCCAAATTTATGTTTTT 3332

Db 62996 ACAAGTGTGATTTGGAAGGCACATGTGAGCATATGCATATATGATCCAAATTTATGTTTTT 62937

QY 3333 CTTTGTATTATATTTTGGGGAAAATTAATAATTTTTTTTAAAGTAAA 3376

Db 62936 CTTTGTATTATATTTTGGGGAAAATTAATAATTTTTTTTAAAGGTATA 62893

RESULT 13

AB037745

LOCUS AB037745 5567 bp mRNA linear PRI 14-MAR-2000

DEFINITION Homo sapiens mRNA for KIAA1324 protein, partial cds.

ACCESSION AB037745

VERSION AB037745.1 GI:7243028

KEYWORDS

SOURCE Homo sapiens brain cDNA to mRNA, clone\_lib:pBluescriptII SK plus clone:fh14139.

ORGANISM Homo sapiens

REFERENCE 1 (sites) Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Nagase,T., Kikuno,R., Ishikawa,K.I., Hirose,M. and Ohara,O.

TITLE Prediction of the coding sequences of unidentified human genes. XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro

JOURNAL DNA Res. 7 (1), 65-73 (2000)

MEDLINE 20181126

REFERENCE 2 (bases 1 to 5567)

AUTHORS Ohara,O., Nagase,T. and Kikuno,R.

TITLE Direct Submission

JOURNAL Submitted (31-JAN-2000) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, URL:http://www.kazusa.or.jp/huge/, Tel: +81-438-52-3913, Fax: +81-438-52-3914)

FEATURES

source Location/Qualifiers

1..5567

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="fh14139"

/tissue\_type="brain"

/clone\_lib="pBluescriptII SK plus"

1..5567

/gene="KIAA1324"

<1..1744

/gene="KIAA1324"

/note="Start codon is not identified."

/codon\_start=2

/product="KIAA1324 protein"

/protein\_id="BAA92562.1"

/db\_xref="GI:7243029"

/translation="WWNTLPNMTTVLSGINFEYKMGTWGEVAGDHIYTAAGASDND FMILITLVPGFRPPQSVNMDTENKEVARITVFETLCSVNCELYFMVGVNSRNTPPVE TWKSKGKQSYTIIIEENTTSFTWAFQRTTFHEASRKYNDVAKIYSINVTNVMNGV ASYCRPCALEASDVSCSPAGYVIDRDSGTCHSCPNTILKAHOPYGVOACVPCG PGTKNNKIHSLYNDCTFSRNPTRTFNFSALANTVTLAGGPSFTSKGLKVEHHFT LSLCNGQGRKMSVCTDNVTDLRIPEGESGFSKITAYVCQAVIIPPEVTGYKAGVSSQ PVSADRLIGVTTDMTLDGITSPAELFHLFLESGLIPDVIFFYRSNDVTQSCSSGRSTTI

BASE COUNT 1382 a 1404 c 1270 g 1511 t

ORIGIN

Query Match 15.3%; Score 517.2; DB 9; Length 5567;

Best Local Similarity 59.0%; Pred. No. 4e-110;

Matches 929; Conservative 0; Mismatches 633; Indels 12; Gaps 2;

QY 1317 ATGGTGAATGTCCCTTCTCTGGCAACATGAAAACTTCCTGCTTCAATGTTGGGAATTCAAA 1376

Db 1 ATGGTGAACACACGCTGCCACAAACATGGAAACGACCGTTCTCAGTGGGATCAACTTCGA 60

QY 1377 GTGCGATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGG 1436

Db 61 GTACAAGGGCATGACAGGCTGGGAGGTGGCTGGTGATCACATTTACACAGCTGCTGGAGC 120

QY 1437 TTCTGACAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAACCACCAACATC 1496

Db 121 CTCAGACAATGACTTCATGATTTCTCACTCTGGTTGTGCCAGGATTTAGACCTCCGCAGTC 180

QY 1497 TATGACTGGAGCCAC---GGGTTCTGAACACTAGGAAGATAACATTTGTCTTTGAGACCCCT 1553

Db 181 GGTGATGGCAGACACAGAGAATAAAGAGGTGGCCAGAAATCACATTTGTCTTTGAGACCCCT 240

QY 1554 CTGTTCACTGACTGTGTTTGTACTTTCATGTTGGGATATTAAAGAAAAAGTACAAATGT 1613

Db 241 CTGTTCTGTGAACCTGTGAGCTCTACTTTCATGGTGGGTGTGAATTTCTAGGACCAACTCC 300

QY 1614 GGTAGAATCGTGGGTGGAACCAAAAGAAAAACAAGCTTACACCCATATCATCTTCAAGAA 1673

Db 301 TGTGGAGACGTGGAAGAGTTTCCAAAGGCAAAACAGTCCCTATACCTACATCATTTGAGGAGAA 360

QY 1674 TGCAACTTTTACATTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGATAAATAGACG 1733

Db 361 CACTACCACGAGCTTCACCTGGCCCTTCCAGAGGACCACCTTTTCATGAGGCAAGCAGGAA 420

QY 1734 GTTCATCAATGACATGGTGAAGATTTATTCTATCACAGCCACTAATGCAAGTTGATGGGGT 1793

Db 421 GTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCAACCAATGTTATGAATGGCGT 480

QY 1794 GCGTCTCATGCCCCGTGCTGCCCTCGGTTCTGAACAGTCGGGTTTCATCGTGTGCC 1853

Db 481 GGCCTCTACTGCGCGTCCCTGTGCCCTAGAACCTCTGTATGTGGGCTCCTCCTGCACCTC 540

QY 1854 CTGCCCTCCAGGCCACTACATTCAGAAAGAAACCAACCCAGTCGAAGGAATGTCCACCTGA 1913

Db 541 TTGTCCTGCTGGTTACTATATTGACCGAGATTTCAGGAACCTGCCACTCCTGCCCCCTAA 600

QY 1914 CACCTACCTGTCCATACATCAGGTCTATGGCAAAGAGGCTTGTATTCATGCGGGCCTGG 1973

Db 601 CACAATTCTGAAAGCCCAACCCAGCCCTTATGGTGTCCAGGCCCTGTGTGCCCTGTGGTCCAGG 660

QY 1974 GAGTAAAAACAATCAGGACCATTCCGTTTGTCTATAGTGACTGCTTTTCTACCATGAAA 2033

Db 661 GACCAAGAACAAAGATCCACTCTCTGTGCTACAATGATTGCACCTTCTCACGCAACAC 720

QY 2034 AGAAATCAGATTTTGCACATATGACTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAA 2093

Db 721 TCCAACCCAGGACTTTCAACTACAACTTCTCCGCTTTGGCAAACACCGTCACTCTTGCTGG 780

QY 2094 TGGCCCCAGCTTCACCTCCAAAGGAACAAAATACTTCCATTTCTCAATATCAGTTTATG 2153

Db 781 AGGGCCAAGCTTCACTTCCAAAGGGTGTGAATACTTCCATCACTTTACCCCTCAGTCTCTG 840

QY 2154 TGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTTTACAGTAAA 2213

Db 841 TGGAAACCAGGGTAGGAAAATGTCTGTGTGACCCGACCAATGTCAGTACCTCCGATTCC 900

QY 2214 AGAAATAGTGGCAGGGTTCAGATGATTACACAAAATTTGGTAGGGGCAATTTGTATGCCAGTC 2273

Db	901	TGAGGGTGAGTCAGGG-----TTCTCCAAATCTATCACAGCCTACGTCTGCCAGGC	951
QY	2274	AACAATATTCCCTTCTGAAGAGTAAGGGTTTCCGAGCAGCCCTTATCATCACAAATCCATCAT	2333
Db	952	AGTCATCATCCCCCAGAGGCTGACAGGCTACAAGGCCGGGGTTTCTTCACAGCCTGTCAG	1011
QY	2334	TCTGGCAGATACATTTCATAGGAGTCACAGTTGAAACCACATTTGAAANAATATTAATATATAA	2393
Db	1012	CCTTGCTGATCGACTTATTGGGGTGACAAACAGATATGACTGTGGATGGAATCACCTCCCC	1071
QY	2394	AGAAGATATGTTCCAGTTTCCAACAGCCCAATACCAGATGTGCATTTCTTTTATAAGTC	2453
Db	1072	AGCTGAACCTTTCCACCTGGAGTCCCTGGGAATACCCGACGTGATCTTCTTTTATAGGTC	1131
QY	2454	TTCTACAGCAACAACATCTTGTATTAATAGCCGATCAACTGCTGTGAAAATGAGGTGTA	2513
Db	1132	CAATGATGTGACCCAGTCCCTGCAAGTTGCTGTCGCCAGGAACGTGCTCAGATGGGACCTG	1191
QY	2514	TCCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCCAGCAAGTGCCCCAGCAGGTACCTG	2573
Db	1192	TCCACAGAAAACCTGTCCCTGGAAAGTTTGCTGTCGCCAGGAACGTGCTCAGATGGGACCTG	1251
QY	2574	TGATGGGTGACGTTCTATTTTCCCTGTGGGAGAGTGCCTGAAGCTTGCCCTCTGTGTACGGA	2633
Db	1252	TGATGGGTGCAACTTCCACTTCCCTGTGGGAGAGCGCGGCTGCTTGCCCGCTCTGCTCAGT	1311
QY	2634	GCATGACTTCCATGAGATTGAGGGAGCCCTGCAAGAGAGAGATTTCAGGAACACCTTGATGT	2693
Db	1312	GGCTGACTACCATGCTATCGTCAAGCAGCTGTGTGGCTGGGATCCAGAAGACTACTTACGT	1371
QY	2694	GTGGAATGAACCTAAATGGTGCATTAAGGAATTTCTTTGCTGAGAAAAAGTTGGCAAC	2753
Db	1372	GTGGCGAGAACCCCAAGCTATGCTCTGGTGGCATTTCTCTGCTGAGCAGAGATCACCAT	1431
QY	2754	CTGTGAACGGTTGACTTTTGGCTGAAGGTGGGAGCGGTGTGGGAGCTTTTACTGCCGT	2813
Db	1432	CTGCAAAACCATAGATTTTCTGGCTGAAGTGGGCATCTCTGCAGGCACCTGTACTGCCAT	1491
QY	2814	TTTGCTGTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATCAAAAAGAAAGAACCAT	2873
Db	1492	CCTGCTCACCGTCTTGACCTGCTACTTTTGGAAAAAGAAATCAAAAACCTAGAGTACAAGTA	1551
QY	2874	TTTGAATCTGTTCA	2887
Db	1552	CTCCAAGCTGGTGA	1565
RESULT 14	AR184096		
LOCUS	AR184096	1129 bp	DNA linear PAT 20-APR-2002
DEFINITION	Sequence 40 from patent US 6342581.		
ACCESSION	AR184096		
VERSION	AR184096.1	GI:20228065	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 1129)		
AUTHORS	Rosen,C.A., Ruben,S.M., Olsen,H.S. and Ebner,R.		
TITLE	Secreted protein HLHP03		
JOURNAL	Patent: US 6342581-A 40 29-JAN-2002;		
FEATURES	Location/Qualifiers		
source	1..1129		
BASE COUNT	379 a 194 c 192 g 363 t	1	others
ORIGIN			
Query Match	14.5%;	Score 493.2;	DB 6; Length 1129;
Best Local Similarity	98.3%;	Pred. No. 1.2e-104;	
Matches	509; Conservative	0; Mismatches	8; Indels 1; Gaps 1;
QY	2860	AAAAAGACACCATTTTGAATCTGTTCAACTGAAAACCTCAAGATCCCCCAATATATGAA	2919

Db	446	AAAAGAAGACCCATTTTGAATCTGTTCAACTGAAAACCTCAAGATCCCCCAATATATGAA	505
QY	2920	GAGACAGTGCCTGTAG-CCTTGAGACTAATGAACAAGAAACCTGCTCTAGTTTACAGGA	2978
Db	506	GAGACAGTGCCTGTAGCCCTTGAGACTAATGAACAAGAAACCTGCTCTAGTTTACAGGA	565
QY	2979	CCATATTTTATAGGCTGTCTCTCATACCTGTACATTTGCTGATCTCACAGAGGAGGCCCAT	3038
Db	566	CCATATTTTATAGGCTGTCTCTCATACCTGTACATTTGCTGATCTCACAGAGGAGGCCCAT	625
QY	3039	GCCGCTGAAAAGGGAGAGATTGAACATTTGATTCCTTATCACATGGTCAAGTACCT	3098
Db	626	GCCGCTGAAAAGGGAGAGATTGAACATTTGATTCCTTATCACATGGTCAAGTACCT	685
QY	3099	TGCCAANAATAAGGAAAGCAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGA	3158
Db	686	TGCCAANAATAAGGAAAGCAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGA	745
QY	3159	AGAGATTTATCTGTATATACACATAACTGAAAACCAAGTTTAAGCCCAACCAATGCACATGC	3218
Db	746	AGAGATTTATCTGTATATACACATAACTGAAAACCAAGTTTAAGCCCAACCAATGCACATGC	805
QY	3219	TGATGCATGCCATATAATTAATGGGTAACTTTTATCTTTATGATGTCTACATAACAAGT	3278
Db	806	TGATGCATGCCATATAATTAATGGGTAACTTTTATCTTTATGATGTCTACATAACAAGT	865
QY	3279	GTGATTTTGGAAAGGCACATGTGAGCATATGCATTTATGATCCAAATTTATGTTTTTCTTTGT	3338
Db	866	GTGATTTTGGAAAGGCACATGTGAGCATATGCATTTATGATCCAAATTTATGTTTTTCTTTGT	925
QY	3339	TTATATATTTTGGGAAAAATTAATAATTTTAAAGGTAAA	3376
Db	926	TTATATTTTGGGAAAAATTAATAATTTTAAAGGTATA	963
RESULT 15	AX336856/c		
LOCUS	AX336856	576 bp	DNA linear PAT 09-JAN-2002
DEFINITION	Sequence 7365 from Patent WO0194629.		
ACCESSION	AX336856		
VERSION	AX336856.1	GI:18127575	
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1		
AUTHORS	Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G., Horrigan,S., Soppet,D.R. and Weaver,Z.		
TITLE	Cancer gene determination and therapeutic screening using signature gene sets		
JOURNAL	Patent: WO 0194629-A 7365 13-DEC-2001;		
FEATURES	Avalon Pharmaceuticals (US)		
source	Location/Qualifiers		
	1..576		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	171 a 112 c 108 g 182 t	3	others
ORIGIN			
Query Match	14.2%;	Score 482.2;	DB 6; Length 576;
Best Local Similarity	97.1%;	Pred. No. 4e-102;	
Matches	501; Conservative	0; Mismatches	14; Indels 1; Gaps 1;
QY	2860	AAAAAGAAGACCATTTTGAATCTGTT-CAACTGAAAACCTCAAGATCCCCCAATATATGA	2918
Db	516	AAAAAGGAGACCCATTTTGAATCTGTTCCAACCTGAACCCCTCAAGATCCCCCAATATATGA	457
QY	2919	AGAGACAGTGTGTAGCCTTGAGACTAATGAACAAAGAAACCTGCTCTAGTTTACAGGA	2978
Db	456	AGAGACAGTGTGTAGCCTTGAGACTAATGANCAAGAAACCTGCTCTAGTTTACAGGA	397
QY	2979	CCATATTTTATAGGGTCTGTCTCTCATACCTGTACATTTGGTGATCTCACAGAGGAGGCCCAT	3038



```

|||||
Db 396 CCATATTTAGGGTCTGTCTCATACCTGTGCACATTGGTGATCTCACAGAGGAGGCCAT 337
QY 3039 GCCGCTGAAAAGGGAGGAGATTGAAACATTTGATTGCCTTATCACATGGTCAAGTACCT 3098
|||||
Db 336 GCCGCTGAAAAGGGAGGAGATTGAAACATTTGATTGCCTTATCACATGGTCAAGTACCT 277
QY 3099 TGCCAAATAAAGGAAAGCAAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGA 3158
|||||
Db 276 TGCCAAATAAAGGAAAGCAAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGA 217
QY 3159 AGAGATTTATCTGTATATACACATAAAGTAAAGCCCAAGTAAAGCCCAAGTAAAGTGC 3218
|||||
Db 216 AGAGATTTATCTGTATATACACATAAAGTAAAGCCCAAGTAAAGCCCAAGTAAAGTGC 157
QY 3219 TGATGCATGCCCATATAATTAATGGGTAACTTTTATTCTTTATGATGTCTACATAACAAGT 3278
|||||
Db 156 TGATGCATGCCCATATAATTAATGGGTAACTTTTATTCTTTATGATGTCTACATAACAAGT 97
QY 3279 GTGATTTGGAAGGCACATGTGAGCATATGCATTATGATCCCAATTTATGTTTTCTTTGT 3338
|||||
Db 96 GTGATTTGGAAGGCACATGTGAGCATATGCATTATGATCCCAATTTATGTTTTCTTTGT 37
QY 3339 TTATATTTTGGGAAAAATTAAAAATTTTAAAGGTA 3374
|||||
Db 36 TTATATTTTGGGAAAAATTAAAAATTTTAAAGGTA 1
```

















QY	2041	CAGATTTTGC	ACTATGA	CTTTAGCA	ACCCTCAGCAGTGTGGGCTCATTAA	TGAATGGCCCC	2100
DB	2041	CAGATTTTGC	ACTATGA	CTTTAGCA	ACCCTCAGCAGTGTGGGCTCATTAA	TGAATGGCCCC	2100
QY	2101	AGCTTCACCT	CCAAAGGA	CAAAATAC	TTCCATTTCTTCA	ATATCAGTTATGTGGGCAT	2160
DB	2101	AGCTTCACCT	CCAAAGGA	CAAAATAC	TTCCATTTCTTCA	ATATCAGTTATGTGGGCAT	2160
QY	2161	GAGGGGAAGA	AGATGGCTCTCTGTACCA	CAATATAACAGACTTTTACAGTAA	AGAAATA	2220	
DB	2161	GAGGGGAAGA	AGATGGCTCTCTGTACCA	CAATATAACAGACTTTTACAGTAA	AGAAATA	2220	
QY	2221	GTGGCAGGGT	CAGATGATTACACAAA	TTTGGTAGGGGCATTTGTATGCCAGTCA	ACAAT	2280	
DB	2221	GTGGCAGGGT	CAGATGATTACACAAA	TTTGGTAGGGGCATTTGTATGCCAGTCA	ACAAT	2280	
QY	2281	ATTTCCTTCT	GAAAGTAAGGGTTCCGAGCAG	CTTATCATCACA	ATCCATCTTGGCA	2340	
DB	2281	ATTTCCTTCT	GAAAGTAAGGGTTCCGAGCAG	CTTATCATCACA	ATCCATCTTGGCA	2340	
QY	2341	GATACATTCA	TAGGAGTCACAGTTGAA	ACCACATTTGAAAA	TATTAATATA	AAAAAGAAGAT	2400
DB	2341	GATACATTCA	TAGGAGTCACAGTTGAA	ACCACATTTGAAAA	TATTAATATA	AAAAAGAAGAT	2400
QY	2401	ATGTTCCCA	GTCCCAACAGCCAAATACC	AGATGTGCATTTCTTTTAT	AAGTCTTCTACA	2460	
DB	2401	ATGTTCCCA	GTCCCAACAGCCAAATACC	AGATGTGCATTTCTTTTAT	AAGTCTTCTACA	2460	
QY	2461	GCAACAACAT	CTTGTATTAA	TGGCGGATCAACTGCTGTGAAAA	TGAGGTGAATCCTACT	2520	
DB	2461	GCAACAACAT	CTTGTATTAA	TGGCGGATCAACTGCTGTGAAAA	TGAGGTGAATCCTACT	2520	
QY	2521	AAATCTGGAG	CAGGAGTGATTTCAGTCCCAGCA	AGTGCCAGCAGGTACCTGTGATGGG	2580		
DB	2521	AAATCTGGAG	CAGGAGTGATTTCAGTCCCAGCA	AGTGCCAGCAGGTACCTGTGATGGG	2580		
QY	2581	TGTACGTTCA	TATTCCTGTGGGAGAGTGCTGA	AGCTTGCCCTCTGTGTACG	GAGCATGAC	2640	
DB	2581	TGTACGTTCA	TATTCCTGTGGGAGAGTGCTGA	AGCTTGCCCTCTGTGTACG	GAGCATGAC	2640	
QY	2641	TTCCATGAGAT	TGAGGGAGCCTGCAAGAGAG	GATTTCAGGAAACCTTTGTATGTGGAAT	2700		
DB	2641	TTCCATGAGAT	TGAGGGAGCCTGCAAGAGAG	GATTTCAGGAAACCTTTGTATGTGGAAT	2700		
QY	2701	GAACCTAAAT	GGTGCAATTAAAGGAATTTCTTGCCT	GAGAAAAAGTTGGCAACCTGTGA	2760		
DB	2701	GAACCTAAAT	GGTGCAATTAAAGGAATTTCTTGCCT	GAGAAAAAGTTGGCAACCTGTGA	2760		
QY	2761	ACGGTTGACT	TTTGGCTGAAGGTGGGAGCCGGTGTGGGAG	CTTTTACTGCCGTTTGTCTG	2820		
DB	2761	ACGGTTGACT	TTTGGCTGAAGGTGGGAGCCGGTGTGGGAG	CTTTTACTGCCGTTTGTCTG	2820		
QY	2821	GTGGCTCTGA	CTGCTACTTCTTGAAAAAAGAAATCA	AAAA	2858		
DB	2821	GTGGCTCTGA	CTGCTACTTCTTGAAAAAAGAAATCA	AAAA	2858		
QY	2859	TTAGTAATGA	CGACTAACTCAAAAGAGTGTGA	ACTCCCGGCTGCAGACAGTTGTGCTATC	2940		
DB	2859	TTAGTAATGA	CGACTAACTCAAAAGAGTGTGA	ACTCCCGGCTGCAGACAGTTGTGCTATC	2940		
QY	2941	ATGGAAGGA	AGAGATAATGAAGAGGAAGTTGTATAT	TCCAATAAACAGTCACTACTAGGA	3000		
DB	2941	ATGGAAGGA	AGAGATAATGAAGAGGAAGTTGTATAT	TCCAATAAACAGTCACTACTAGGA	3000		
QY	2859	-----	-----	-----	-----	2858	
DB	2859	-----	-----	-----	-----	2858	
QY	2895	ACCTCAAGAT	CCCCAAATATATGAAGAGACAGT	GTGCTGTAGCCTTGAGACTAATGA	ACAAA	2954	
DB	3061	ACCTCAAGAT	CCCCAAATATATGAAGAGACAGT	GTGCTGTAGCCTTGAGACTAATGA	ACAAA	3120	

QY	2955	GAACACCTGCTCTAGTTTTACAGGACCATATTTTAGGGTCTGTCTCTCATACCTGTCACATT	3011
DB	3121	GAACACCTGCTCTAGTTTTACAGGACCATATTTTAGGGTCTGTCTCTCATACCTGTCACATT	3180
QY	3015	GGTGATCTCACAGAGAGGGCCCATGCCGCTGAAAAGGGAAGGAGATTGAAACATTTGATT	3074
DB	3181	GGTGATCTCACAGAGAGGGCCCATGCCGCTGAAAAGGGAAGGAGATTGAAACATTTGATT	3240
QY	3075	GCCTTATCACATGGTCAAGTACCTTGCCCAATAAAGGAAAGCAAATGATTGGGTCTCAA	3134
DB	3241	GCCTTATCACATGGTCAAGTACCTTGCCCAATAAAGGAAAGCAAATGATTGGGTCTCAA	3300
QY	3135	CTGAAGATGAAGCTCAACTCAGGAAGAGATTATCTGTATATACACATACTGAAACCA	3194
DB	3301	CTGAAGATGAAGCTCAACTCAGGAAGAGATTATCTGTATATACACATACTGAAACCA	3360
QY	3195	AGTTTAAGCCCAATGCACCTGCTGATGCATGCCATATAATTAATGGGTAACCTTTATT	3254
DB	3361	AGTTTAAGCCCAATGCACCTGCTGATGCATGCCATATAATTAATGGGTAACCTTTATT	3420
QY	3255	CTTTATGATGTCTACATAACCAAGTGTGATTGGAGGACACATGTGAGCATATGCATTATG	3314
DB	3421	CTTTATGATGTCTACATAACCAAGTGTGATTGGAGGACACATGTGAGCATATGCATTATG	3480
QY	3315	ATCCAATTTATGTTTTTCTTTGTTTATATTTTGGGGAAAAATTAAATTTTTTAAGGTA	3374
DB	3481	ATCCAATTTATGTTTTTCTTTGTTTATATTTTGGGGAAAAATTAAATTTTTTAAGGTA	3540
QY	3375	AAAAAAAAAAAAAAAA 3390	
DB	3541	AAAAAAAAAAAAAAAA 3556	
RESULT 3			
AAC84891			
ID	AAC84891 standard; cDNA; 1737 BP.		
XX			
AC	AAC84891;		
XX			
DT	20-APR-2001 (first entry)		
XX			
DE	Human SEC10 nucleic acid sequence (clone ID 1795045.0.77).		
XX			
KW	SECX; cytostatic; gynecological; gene therapy; screening assay; human;		
KW	chromosomal mapping; forensic biology; cell proliferation; cancer;		
KW	cell differentiation; immune associated disorder; gestational disease;		
KW	SEC10; ss.		
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	296..1690	
FT	/*tag= a		
FT	/product= "SEC10"		
XX			
PN	WO200078802-A2.		
XX			
PD	28-DEC-2000.		
XX			
PF	23-JUN-2000; 2000WO-US17328.		
XX			
PR	23-JUN-1999; 99US-0140584.		
PR	20-JUL-1999; 99US-0144722.		
PR	16-SEP-1999; 99US-0154520.		
PR	22-JUN-2000; 2000US-0604286.		
XX			
PA	(CURA-) CURAGEN CORP.		
XX			
PI	Shimkets RA, Fernandes E, Vernet C, Yang M, Boldog FL;		
PI	Herrmann JL;		
XX			
DR	WPI; 2001-071385/08.		
DR	P-PSDB; AAB48377.		







DE Human CASB619 protein coding sequence #1.  
XX  
KW Human; CASB619; cancer; autoimmune disease; immunogen; vaccine;  
KW epitope; ss.

OS Homo sapiens.

	Key	Location/Qualifiers
FT	CDS	1..3342
FT		/*tag= a
FT		/product= "CASB619"

PN WO200058460-A2.

PD 05-OCT-2000.

PF 20-MAR-2000; 2000WO-EP02478.

PR 26-MAR-1999; 99GB-0007113.

PR 25-SEP-1999; 99GB-0022858.

PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

PI Bruck CEM, Cassart J, Coche T, Vinals De Bassols YC;

DR WPI; 2000-664923/64.

DR P-PSDB; AAB26179.

PT Novel CASB619 polypeptides useful for diagnosing, and as vaccines for  
PT prophylactic and therapeutic treatment of, cancers, particularly  
PT ovarian and colon carcinoma, and autoimmune diseases -

PS Claim 13; Page 53-54; 68pp; English.

The present sequence comprises the human CASB619 coding sequence. This protein is thought to be specifically or over-expressed in tumour cells, and so can be used as a target for antigen-specific immune responses which can cause destruction of the tumour cell. In addition, the protein and gene can be used in cancer diagnosis, in the treatment of autoimmune diseases and in vaccines against cancer and autoimmune disease. The invention provides a number of epitopes derived from the protein which can be used as immunogens.

SQ Sequence 3280 BP; 810 A; 930 C; 797 G; 743 T; 0 other;

Query Match 26.7%; Score 904.6; DB 21; Length 3280;  
Best Local Similarity 60.1%; Pred. NO. 5.2e-27;  
Matches 1637; Conservative 0; Mismatches 1059; Indels 27; Gaps 7;

QY 171 GCTTCCTCCTGCCAGGAGAAAGATTATCACCCTTTGAATATACGGAAATGTGATAGCAGTGG 230

Db 135 GCTTCATGCCCTGCAAAGAGTCTGAGTACCCACTATGAGTACACGGCGTGTGACAGCACGGG 194

QY 231 CTCCAGGTGGAGAGTTGCCATTCCAAATTCTGCAGTGGACTGCTCTGGCCCTGCCCTGACCC 290

Db 195 TTCCAGGTGGAGGGTCCCGTGCCGCATACCCCGGGCCCTGTGCACCCAGCCTGCCCTGACCC 254

QY 291 AGTGAGAGGCAAGAAATGCACCTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAA 350

Db 255 CGTCAAGGGCACCGAGTGCTCCTTCTCCTGCAACGCCGGGGAGTTTCTGGATATGAAGGA 314

QY 351 CCAGGTATGCAGTAAGTGTGGTGAAGGCACCTATTCCCTTGGGCAGTGGCATCAAAATTGA 410

Db 315 CCAGTCATGTAAGCCATGCGCTGAGGGCCGCTACTCCCTCGGCACAGGCATTTCGGTTTGA 374

QY 411 TGAATGGGATGAATTGCCGGCAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGT 470

Db 375 TGAGTGGGATGAGCTGCCCCATGGCTTTGCCAGCCCTCTCAGCCAACATGGAGCTGGATGA 434

QY 471 GGGCCCTTCTGACAGCAGGCCAGCGCTGTAAACAACCTCTTCTGGATCCCTCGTGGAAA 530

Dbb 435 CAGTGGCTGAGTCCA--CCGGAACTGTA<sup>T</sup>CTCGTCCAAGTGGGTCCCCGGGGCGGA 491

Db 1563 CAACACTCCTGTGGAGACGTGGAAGGTTCCAAAGGCAAAACAGTCTTATACCTACATCAT 1622  
QY 1665 CTTCAGAATGCAACTTTTACATTTACATGGGCATTCCAGAGAACTAATCAGGGTCAAGA 1724  
Db 1623 TGAGGAGAACACTACCACGAGCTTCACCTGGCCCTCCAGAGGACCACTTTTCATGAGGC 1682  
QY 1725 TAATAGACGGTTTCATCAATGACATGGTGAAGATTTATTCTATCAGACCCACTAATGCAGT 1784  
Db 1683 AAGCAGGAAGTACACCAATGACGTGCCCAAGATCTACTCCATCAATGTCACCAATGTTAT 1742  
QY 1785 TGATGGGGTGGCGTCCCTCATGCGGTGCCCTGTGCCCTCGGTTCTGAACAGTCGGGTTTCATC 1844  
Db 1743 GAATGGCGTGGCCTCCCTACTGCGGTGCCCTGTGCCCTAGAACGCTCTGATGTGGGCTCCTC 1802  
QY 1845 GTGTGTCCTCCCTGCCCTCCAGGCCACTACATTGAGAAAGAAACCAACCAAGTGAAGGAATG 1904  
Db 1803 CTGCACCTCTGTCTCCTGTGTTACTATATTGACCGAGATTACAGGAACCTGCCACTCCTG 1862  
QY 1905 TCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTTATTCCTATG 1964  
Db 1863 CCCCCCTAACACAATTTCTGAAGCCCAACAGCCTTATGGTGTCCAGGCCCTGTGTGCCCTG 1922  
QY 1965 CGGGCCTGGGAGTAAAAACAATCAGGACCATTTCGGTTTGTCTATAGTACTGCTTTTCTA 2024  
Db 1923 TGGTCCAGGGACCAAGAACAACAGATCCACTCTCTGTGTCTACAATGATTGCACCTTCTC 1982  
QY 2025 CCATGAAAGAAATCAGATTTTGCACACTATGACTTTAGCAACCTCAGCAGTGTGGGCTC 2084  
Db 1983 ACGCAACACTCCAACACGAGACTTTCAACTACAACCTCTCCGCTTTGGCAAAACCCGTAC 2042  
QY 2085 ATTAATGAATGGCCCCAGCTTCACTCCAAAGGAACAAATFACTTCCATTTCTTCAATAT 2144  
Db 2043 TCTTGTGGAGGGCCCAAGCTTCACTTCCAAAGGGTTGAAATACTTCCATCACTTTACCT 2102  
QY 2145 CAGTTTATGTGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTT 2204  
Db 2103 CAGTCTCTGTGGAACCCAGGTAGGAAATGTCTGTGTGCACCGACAATGTCACTGACCT 2162  
QY 2205 TACAGTAAAGAAATAGTGGCAGGTCAGATGATTACACAAATTTGGTAGGGGCATTTGT 2264  
Db 2163 CCGGATTCCTGAGGGTGAGTCAGG-----TTCTCCAAATCTATCACAGCCTACGT 2213  
QY 2265 ATGCCAGTCAACAATATTATTCCTTCTGAAAGTAAGGGTTTCCGAGCAGCCTTATCATCACA 2324  
Db 2214 CTGCCAGGCAGTCATCATCCCCCAGAGGTGACAGGCTACAGGCCGGGGTTTCCTCACA 2273  
QY 2325 ATCCATCATTTCTGGCAGATACATTTCATAGGAGTCACAGTTGAACCCACATTGAAAAATAT 2384  
Db 2274 GCCTGTAGCCTTGCTGATCGACTTATTGGGGTGACAACAGATATGACTCTGGATGGAAT 2333  
QY 2385 TAATATAAAGAAGATATGTTCCAGTTTCCAAACAAGCCAAATACCAGATGTGCATTTCTT 2444  
Db 2334 CACCTCCCAGCTGAACCTTTTCCACCTGGAGTCTCTGGGAATACCGGACGTGATCTTCTT 2393  
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTTAATGGCCGATCAACTGCTGTGAAAAAT 2504  
Db 2394 TTATAGGTCCAATGATGTGACCCAGTCTCTGCAGTTCTGGGAGATCAACCACTCCGCGT 2453  
QY 2505 GAGGTGTAATCCTACTAAATCTGAGCAGGAGTGATTTTCAGTCCCCAGCAAGTGCCTCAGC 2564  
Db 2454 CAGGTGCAGTCCACAGAAAACTGTCCCTGGAAGTTTGCTGCTGCCAGGAACGTGCTCAGA 2513  
QY 2565 AGGTACCTGTGATGGGTGTACGTTCTATTTCCTGTGGGACAGTGTGAAGCTTGCCCTCT 2624  
Db 2514 TGGGACCTGTGATGGCTGCAACTTCCACTTCCTGTGGGAGAGCGCGCTGCTTGCCCGCT 2573  
QY 2625 GTGTACGGAGCATGACTTCCATGAGATTTGAGGAGCCTGCAAGAGAGGATTTTCAGGAAC 2684  
Db 2574 CTGCTCAGTGGCTGACTACCATGCTATCGTCAGCAGCTGTGTGGTGGGATCCAGAGAC 2633  
QY 2685 CTTGTATGTGTGGAATGAACCTAAATGGTGCATTAAGGAATTTCTTTGCCTGAGAAAAA 2744

Db 2634 TACTTACGTGTGGCGAGAACCCCAAGCTATGCTCTGGTGGCATTTTCTCTGCCTGAGCAGAG 2693  
QY 2745 GTTGGCAACCTGTGAACCGTTGACTTTTGGCTGAAGGTGGAGCGGTTGTGGAGCTTT 2804  
Db 2694 AGTCACCATCTGCAAAACCATAGATTTTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG 2753  
QY 2805 TACTGCGTTTTGTGCTGGTCTGACCTGCTACTTCTGGAAAAAGAAATCAAAAGAAAA 2864  
Db 2754 TACTGCCATCTGCTCACCGTCTTGACCTGCTACTTTTGGAAAAAGAAATCAAAACTAGA 2813  
QY 2865 GAAGACCATTTTGAATCTGTTCA 2887  
Db 2814 GTACAAGTACTCCAAGCTGGTGA 2836  
RESULT 6  
AAF89765  
ID AAF89765 standard; cDNA; 3331 BP.  
XX  
AC AAF89765;  
XX  
DT 23-JUL-2001 (first entry)  
XX  
DE Nucleotide sequence of a human protein expressed in tumour cells.  
XX  
KW Tumour cell; immunological disease; autoimmune disease; cancer;  
XX  
OS infection; ss.  
XX  
OS Homo sapiens.  
FH Key Location/Qualifiers  
CDS 45..3086  
FT /\*tag= a  
FT /product= "tumour expressed protein"  
FT sig\_peptide 45..167  
FT /\*tag= b  
XX  
PN WO200131003-A1.  
XX  
PD 03-MAY-2001.  
XX  
PF 30-OCT-2000; 2000WO-FR03032.  
XX  
PR 29-OCT-1999; 99FR-0013629.  
XX  
PA (FABR ) FABRE MEDICAMENT SA PIERRE.  
XX  
PI Delneste Y, Magistrelli G, Jeannin P, Bonnefoy J;  
XX  
DR WPI; 2001-328651/34.  
DR P-PSDB; AAB83845.  
XX  
PT New nucleic acid, expressed in tumours and lymphoid tissue is useful for  
XX identifying agents for treating tumours and autoimmune disease  
PS Claim 1; Page 43-48; 85pp; French.  
XX  
CC The present sequence encodes a human protein expressed in tumour cells.  
CC The polynucleotide is useful for screening cDNA/genomic DNA banks and  
CC for cloning isolated DNA; identifying mutant forms of the gene that  
CC encodes a human protein, where the mutations are associated with  
CC abnormal gene expression, or promoters and regulators of the gene,  
CC particularly for diagnosis; for recombinant expression of the derived  
CC protein; as probes and primers for detection and amplification; and  
CC as antisense therapeutics. The tumour expressed protein is useful for  
CC raising specific antibodies and to screen agents that modulate its  
CC activity, bind to it or interact with it. These agents are potentially  
CC useful for treatment or prevention of diseases associated with abnormal  
CC expression/activity of the protein, particularly immunological diseases  
CC (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic  
XX infections.  
SQ Sequence 3331 BP; 822 A; 945 C; 813 G; 751 T; 0 other;

Query Match		26.6%	Score 903;	DB 22;	Length 3331;
Best Local Similarity		60.1%	Pred. No. 1.4e-226;		
Matches 1636; Conservative		0;	Mismatches 1060;	Indels 27;	Gaps 7;
Qy	171	GCTTCCTCCTGCCAGGAGAAAGATTATACACTTTTGAATATACGGAATGTGATAGCAGTGG	230		
Db	179	GCTTCATGCTGCAAGAGTCTGAGTACCACATATGAGTACACGCGGTGTGACAGCACGGG	238		
Qy	231	CTCCAGGTGGAGAGTTGCCATTCCCAATTTCTGCAGTGGACTGCTCTGGCCCTGCCTGACCC	290		
Db	239	TTCCAGGTGGAGGGTCGCCGTGCCGCATACCCCGGCCCTGTGCAACAGCCTGCCTGACCC	298		
Qy	291	AGTGAGAGCGCAAGAATGCACATTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAA	350		
Db	299	CGTCAAGGCGACCGAGTGCTCCTCTCTGCAACGCCGGGAGTTCTTGGATATGAAGGA	358		
Qy	351	CCAGGTATGCAAGTAAGTGTGGTGAAGGACCACTATTCTTGGGAGTGGCATCAAAATTTGA	410		
Db	359	CCAGTCATGTAAAGCCATGCGCTGAGGCGCGCTACTCCCTCGGCACAGGCAATTCGGTTTGA	418		
Qy	411	TGAATGGGATGAATTGCCTGGCAGGATTTTCTAAACATCGCAACATTCATGGACACTGTGGT	470		
Db	419	TGAGTGGGATGAGCTGCCCATGGCTTTGCCAGCCTCTCAGCCCAACATGGAGTGGATGA	478		
Qy	471	GGGCCCTCTGCACAGCAGGCCGAGCGCTGTAAACAACTCTTCTTGGATCCTCTCGTGGAAA	530		
Db	479	CAGTGTCTGAGTGCA--CCGGGAAGTGTACTTCGTCCAAGTGGGTTCCTCCGGGGCGA	535		
Qy	531	CTACATAGAAATCTAATCGTGATGACTGCACGGGTGTCTTTTGATCTATCTGCTGACACCTAA	590		
Db	536	CTACATCGCTCCCAACACGGACGGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA	595		
Qy	591	GAAGTCAGGCTATGTCTTCTTTGAGTACCAGTATGTGCGACAACACACATCTTCTTTGAGTT	650		
Db	596	GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT	655		
Qy	651	CTTTATTCAAAATGATCAGTGCCAGGAGATGGACACCACCTGACAAGTGGGTAAAACT	710		
Db	656	TTTCGTTCAAGATGACCAGTGCCAGGCCCAATGCAGATGACTC--CAGGTGGATGAAGAC	712		
Qy	711	TACAGACAATGGAGAATGGGGCTCTCATFTCTGTAATGCTGAATTCAGGCACAAACATACT	770		
Db	713	CACAGAGAA--AGGATGGGAATTCACAGTGTGGAGCTAAATCGAGGCAATAATGTCTT	769		
Qy	771	CTACTGGAGAACTACAGGCATCCTTATGGGTCTTAAGCGGTCAAGCCTGTGCTGGTAAA	830		
Db	770	CTATTGGAGAACACAGCCTTCTCAGTATGGACCACAAAGTACCCAAAGCCTGTGCTGGTGAG	829		
Qy	831	AAATATCACAATTAAGGGGTGGCTACACATCAGAATGTTTTCTTCCAGCAAGCCAGGCAC	890		
Db	830	AAACATTTGCCATAACAGGGGTGGCTACACTTCAGAATGCTTCCCTTCCCTGCAAACTGGCAC	889		
Qy	891	ATTTCAGCAACAACACAGGTTTCATTTCAACTGCCAGGTGTGTCCAGAAACACCTATTCTGA	950		
Db	890	GTATGCAGACAAGCAGGGCTCCTCTTTCTGCAAACTTTGCCAGCAACTCTTATTTCAAA	949		
Qy	951	GAAAGGAGCCAAAGAATGTATAAGGTGTAAA--GACGACTCTCAATTTTCAGGATCCAG	1007		
Db	950	TAAAGGAGAAACTTCTTGCCACCAGTGTGACCCTTGACAAATACTCAGAGAAAGGATCTTC	1009		
Qy	1008	TGAGTGTACAGAGCGCCCTCCCTGTACACACAAAGACTATTTCCAGATCCATACTCCATG	1067		
Db	1010	TTCCCTGTAACTGCGCCAGCTTGCACAGACAAAGATTATTTCTACACACACACGCGCTG	1069		
Qy	1068	TGATGAAGAGGAAAGACACAGATAATGTACAAGTGGATAGAGGCCCAAAATCTGCCGGGA	1127		
Db	1070	CGATGCCAACGGAGACACAACTCATGTACAAATGGGCCAAGCCGAAATCTGTAGCGA	1129		
Qy	1128	GGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAGAGAGGATTTGTCGGCCTTG	1187		
Db	1130	GGACCTTGAGGGGGCAGTGAAGCTGCCTGCCTCTGGTGTGAAGAGCCCACTGCCCAACCTG	1189		

Qy	1188	CAACCCCTGGATTTTATAAACAATGGATCATCTTCTTGCCATCCCTGTCTCCTGGAACATTT	1247		
Db	1190	CAACCCAGGCTTCTTCAAAACCACACAGCACCTGCCAGCCCTGCCATATGGTCCCTA	1249		
Qy	1248	TTCAGATGGAACCAAAAGAAATGTAGACCATGTCCAGCAGGAACGGAGCTGCACTTGGCTT	1307		
Db	1250	CTCCAATGG--CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGATTT	1306		
Qy	1308	TGAATATAAATGGTGGAAATGTCTTCTGGCAACATGAAAACCTTCTGCTTCAATGTTGG	1367		
Db	1307	TGAATACAAATGGTGAACACGCTGCCACACAAACATGGAACACGACCGTCTCAGTGGGAT	1366		
Qy	1368	GAATTCAAAGTCCGATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGG	1427		
Db	1367	CAACTTCGAGTACAAGGCAATGACAGCTGGGAGGTGGCTGGTGTACATTTACACAGC	1426		
Qy	1428	GGCTGGAGGTTCTGACAAATGATTTACCTGATCTTAAACTTGCATATCCAGGATTTAAACC	1487		
Db	1427	TGCTGGAGCCTCAGACAAATGACTTCATGATTTCTCACTCTGGTTGTGCCAGGATTTAGACC	1486		
Qy	1488	ACCAACATCTATGACTGGAGCCAC--GGGTTCTGAACCTAGGAAGAAATAACATTTGTCTT	1544		
Db	1487	TCCGCGAGTCGGTGATGGCAGACACAGAGAAATAAGAGGTGGCCAGAAATCATATTTGTCTT	1546		
Qy	1545	TGAGACCCCTCTGTTCAGCTGACTGTGTTTGTACTTTCATTTGATGGTGGATTAATAAGAAAAG	1604		
Db	1547	TGAGACCCCTCTGTCTGTGAACCTGTGAGCTTACTTTCATTTGTTGGTGGTGTGAATTTCTAGGAC	1606		
Qy	1605	TACAAATGTGGTAGAATCGTGGGTGGAAACCAACAAAGAAACAAAGCTTACACCCATATCAT	1664		
Db	1607	CAACACTCCTGTGGAGACGTGGAAAGGTTCCAAANGGCAACAGTCCCTATACCTTACATCAT	1666		
Qy	1665	CTTCAAGAAATGCAACTTTTACATTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGA	1724		
Db	1667	TGAGGAGAAACACTACCACGAGCTTCACCTGGGCCCTTCCAGAGGACCACCTTTTCATGAGGC	1726		
Qy	1725	TAATAGACGCTTCATCAATGACATGGTGAAGATTTATTTCTATCACAGCCACTAATGCAGT	1784		
Db	1727	AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCAACCAATGTTAT	1786		
Qy	1785	TGATGGGTGGCTCCTCATGCGCTGCTGTGCCCTCGGTTCTGACAGTGGGTTCATC	1844		
Db	1787	GAATGGCGTGGCTCCTACTTGCCTGCTGTGCCCTTGAAGCCCTTGATGTGGCTTCTCTC	1846		
Qy	1845	GTGTGTCCCTGCCCTCCAGGCCACTACATTCAGGTCTATGGCAAAAGAGGCTGTATTTCCATG	1904		
Db	1847	CTGCACCTCTTGTCTGTGTTACTATATATGACCGAGATTCAGGAACCTGCCACTTCTCTG	1906		
Qy	1905	TCCACCTGACACCTTACCTGTCCATACATCAGGTCTATGGCAAAAGAGGCTGTATTTCCATG	1964		
Db	1907	CCCCCTTAACACAAATCTGAAAGCCCCACAGCCTTATGGTGTCCAGGCTGTGTGCCCTG	1966		
Qy	1965	CGGGCTTGGAGTAAACAAATCAGGACCATTCCGTTTGTGCTATAGTACTGCTTTTCTA	2024		
Db	1967	TGGTCCAGGACCAAGAACAACAGATCCACTCTCTGTGCTACAAATGATTCACCTTCTC	2026		
Qy	2025	CCATGAAAAGAAAATCAGATTTTGCACATATGACTTTTAGCAACCTCAGCAGTGTGGGCTC	2084		
Db	2027	ACGCAACACTCCCAACCAAGGACTTTCAACTACAACTTCTCGGCTTTGGCAACACCGTCAAC	2086		
Qy	2085	ATTAAATGAATGGCCCCCAGCTTCCACCTCCAAAGGAACAAAATATCTTCCATTTCTCAATAT	2144		
Db	2087	TCTTGTGGAGGGCCNAGCTTCACCTTCCAAAGGGTTGAAATATCTTCCATCCTTTACCTT	2146		
Qy	2145	CAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTATACCAACAAATATAACAGACTT	2204		
Db	2147	CAGTCTCTGTGGAAACCAGGGTAGGAAAATGTCTGTGTGCACCGACAATGTCACTGACCT	2206		
Qy	2205	TACAGTAAAGAAAATAGTGGCAGGGTACAGATGATTAACAAAAATTTGGTAGGGCATTTGT	2264		
Db	2207	CCGGATTCCTGAGGGTGTGATCAGGG-----TTCTCCAAATCTATCACAGCCTACGT	2257		
Qy	2265	ATGCCAGTCAACAATTTATTTCTTCTGAAAAGTAAAGGTTTCCGAGCAGCCTTATCATATCACA	2324		



Db 2258 CTGCCAGGCAGTCATCATCCCCCAGAGAGGTGACAGGCTACAAGGCCGGGGTTCTCTCACA 2317  
QY 2325 ATCCATCATCTCTGGCAGATACATTCATAGGAGTCACAGTTGAAACCACATTTGAAAAATAT 2384  
Db 2318 GCCTGTAGCCCTTGTGATCGACTTATTGGGGTGACAAACAGATATGACTCTGGATGGAAT 2377  
QY 2385 TAATATAAAGAAGATATGTTCCAGTTCCAAACAGCCAAATACCAGATGTGCATTTCTT 2444  
Db 2378 CACCTCCCAGCTGAACCTTTTCCACCTGGAGTCTCTGGGAATACCCGACGTGATCTTCTT 2437  
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAAATGGCCGATCAACTGCTGTGAAAAT 2504  
Db 2438 TTATAGGTCCAATGATGTGACCCAGTCTCTGCAGTTCTGGGAGATCAACCCACCTCCGCT 2497  
QY 2505 GAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCCAGCAAGTCCCCAGC 2564  
Db 2498 CAGGTGCAGTCCACAGAAAACGTCTCCCTGGAAGTTGCTGCTGCCAGGACGTGCTCAGA 2557  
QY 2565 AGGTACCTGTGATGGGTGACGTTCTATTATTTCCCTGTGGGAGAGTGCTGAAGCTTGCCCTCT 2624  
Db 2558 TGGGACCTGTGATGGCTGCAACTTCCACTTCCTGTGGGAGAGCGCGCTGCTTGCCCGCT 2617  
QY 2625 GTGTACGGAGCATGACTTCCATGAGATTGAGGGGAGCCTGCAAGAGAGAGGATTCAGGAAAC 2684  
Db 2618 CTGCTCAGTGGCTGACTACCGTGTCTATCTGTCAGCAGCTGTGTGGTGGGATCCAGAAGAC 2677  
QY 2685 CTTGTATGTGTGGAATGAACCTAAATGGTGCAATTAAGGAATTTCTTTGCCCTGAGAAAAA 2744  
Db 2678 TACTTACGTGTGGCGAGAACCCCAAGCTATGCTCTGGTGGCATTTCTCTGCCCTGAGCAGAG 2737  
QY 2745 GTTGGCAACCTGTGAACGGTTGACTTTTGGCTGAAGGTGGGAGCGGCTGTGGGAGCTTT 2804  
Db 2738 AGTCACCATCTGCAAAACCATAGATTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG 2797  
QY 2805 TACTGCCGTTTGTGCTGGTCTGTGACCTGTACTTCTGGAAAAAGAAATCAAAAGAAAAA 2864  
Db 2798 TACTGCCATCTGCTCACCGTCTTGACCTGTACTTCTTGAAAAAGAAATCAAAACTAGA 2857  
QY 2865 GAAGACCATTTTGAATCTGTTCA 2887  
Db 2858 GTACAAGTACTCCAAGCTGGTGA 2880

RESULT 7  
AAF28030  
ID AAF28030 standard; DNA; 3334 BP.  
XX  
AC AAF28030;  
XX  
DT 08-MAY-2001 (first entry)  
XX  
DE Human TR13 receptor coding sequence SEQ ID NO: 39.  
XX  
KW Human; tumour necrosis factor receptor; TR13; TR14; infection;  
KW cancer; autoimmune disease; allergy; inflammatory disease;  
KW graft rejection; apoptosis; cardiovascular disease; aneurysm; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200105834-A1.  
XX  
PD 25-JAN-2001.  
XX  
PF 14-JUL-2000; 2000WO-US19343.  
XX  
PR 16-JUL-1999; 99US-0144087.  
PR 18-AUG-1999; 99US-0149450.  
PR 20-AUG-1999; 99US-0149712.  
PR 10-SEP-1999; 99US-0153089.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX

PI Ruben SM, Ni J, Young PE;  
XX WPI; 2001-112682/12.  
DR  
XX Nucleic acids encoding 2 human tumor necrosis factor receptor  
PT polypeptides ((TR13) and (TR14)), useful for the prevention, diagnosis  
PT and treatment of, e.g. cancers, acquired immune deficiency syndrome and  
PT hypohidrotic ectodermal dysplasia -  
XX  
PS Claim 4; Page 394-398; 418pp; English.  
XX  
CC The present invention provides the protein and coding sequences of the  
CC human tumour necrosis factor receptors TR13 and TR14. These sequences are  
CC useful in the diagnosis and treatment of many diseases, including cancer,  
CC autoimmune diseases, cardiovascular disorders, allergies,  
CC neurodegenerative diseases, graft rejection, inflammation, aneurysms and  
CC infections.  
XX  
SQ Sequence 3334 BP; 820 A; 952 C; 811 G; 751 T; 0 other;  
  
Query Match 26.6%; Score 901.4; DB 22; Length 3334;  
Best Local Similarity 60.0%; Pred. No. 3.7e-226;  
Matches 1635; Conservative 0; Mismatches 1061; Indels 27; Gaps 7;  
  
QY 171 GCTTCCTCCTTGCCAGAGAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGG 230  
Db  
QY 192 GCTTCACGCCCTGCAAGAGTCTGAGTACCACACTATGAGTACACGGCGTGTGACAGACGGG 251  
Db  
QY 231 CTCCAGGTGGAGAGTTGCCATTCCAAATTTCTGCAGTGGACTGCTCTGGCCTGCCTGACCC 290  
Db  
QY 252 TTCCAGGTGGAGGGTGCCTGCTGCCGATACCCCGGCTGTGCACACGCTGCCTGACCC 311  
QY 291 AGTGAGAGGCAAGAAATGCATTTCTCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAA 350  
Db  
QY 312 CGTCAAGGGCACCGAGTCTCTCTCTGCAACGCCGGGAGTTTCTGGATATGAAGGA 371  
QY 351 CCAGGTATGCAGTAAGTGTGGTGAAGGCACCTATTCTTGGCAGTGGCAGTCAAAATTTGA 410  
Db  
QY 372 CCAGTCATGTAAGCCATCGCTGAGGGCCGCTACTCCCTGGCACAGGCATTCGGTTGA 431  
QY 411 TGAATGGGATGAATTCGCCGCAGAGTATTTCTAACATCGCAACATTCATGGACACTGTGGT 470  
Db  
QY 432 TGAGTGGGATGAGTGCCTCCATGGCTTTGCCAGCCTCTCAGCCAAACATGGAGCTGGATGA 491  
QY 471 GGGCCCTTCTGACAGCAGGCCAGACGGCTGTAAACAACCTCTTCTTGGATCCCTCGTGA 530  
Db  
QY 492 CAGTGTCTGTAGTCCA--CCGGAACTGTACTTCTGTCCTCAAGTGGGTCCCCCGGGCGGA 548  
QY 531 CTACATAGAAATCTAATCGTGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAA 590  
Db  
QY 549 CTACATCGCCTTCAACACGACGACGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 608  
QY 591 GAAGTCAGGCTATGTCTTCTTTGAGTACCAGTATGTCGACAAACAACATCTTCTTTGAGTT 650  
Db  
QY 609 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 668  
QY 651 CTTTATTCAAATGATCAGTCCAGGAGATGGACACCACTGACAAAGTGGGTAAACT 710  
Db  
QY 669 TTTCGTTCAGAAATGACCAAGTCCAGCCCAATGCAGATGACTC--CAGGTGGATGAAGAC 725  
QY 711 TACAGACAATGGAGAATGGGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATACT 770  
Db  
QY 726 CACAGAGAA--AGGATGGAAATTCACAGTGTGGAGCTAAATCGAGGCAATAATGTCTT 782  
QY 771 CTACTGGAGAACTACAGGCATCCTTATGGGTCTTAAGCGGTCAAGCCTGTGCTGGTAAA 830  
Db  
QY 783 CTATTGGAGAACCCACAGCCTTCTCAGTATGGACCAAAAGTACCCAAAGCCTGTGCTGGTGAG 842  
QY 831 AAATATCACAAATTGAAGGGTGGCGGTACACATCAGAAATGTTTCTTGTCAAGCCAGGCAC 890  
Db  
QY 843 AAACATTGCCATAACAGGGGTGGCCTACACTTCAGAAATGCTTCCCTGCCAAACCTGGCAC 902  
QY 891 ATTCAGCAACAACACCAGGTTTCATTCAACTGCCAGGTGTGTCCAGAAACACCTATTCTCTGA 950

Db 903 GTATGCAGACAAGCAGGGCTCCTCTTTCTGCAAACTTTGCCCAGCCAACCTCTTATTTCAAA 962  
QY 951 GAAAGGAGCCAAAGAATGTATAAGGTGTAAA--GAGCACTCTCAATTTTCAGGATCCAG 1007  
Db 963 TAAAGGAGAAACTTCTTGCCACCAGTGTGACCCTGACAAATACTCAGAGAAAGGATCTTC 1022  
QY 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCAGATCCATACTCCATG 1067  
Db 1023 TTCTGTAAAGTGGCCCAAGCTTGCACAGACAAAGATTATTCTACACACACACGGCCTG 1082  
QY 1068 TGATGAAGAAGGAAAGACACAGATAAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGGA 1127  
Db 1083 CGATGCCAACGGAGAGACACAACACTCATGTACAAATGGGCCAAGCCGAAATCTGTAGCGA 1142  
QY 1128 GGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAAGAAGGATTGTCCGCCTTG 1187  
Db 1143 GGACCTTGAGGGGGCAGTGAAGCTGCCTGCCTCTGGTGTGAAGACCCACTGCCACCCCTG 1202  
QY 1188 CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCCTGGAACATT 1247  
Db 1203 CAACCCAGGCTTCTTCAAAACCAACAACAGCACCTGGCCAGCCCTGCCCATATGGFTTCTA 1262  
QY 1248 TTCAGATGGAACCAAGAAATGTAGACCATGTCCAGAGGAACGGAGCCTGCACCTTGGCTT 1307  
Db 1263 CTCCAATGG--CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGATT 1319  
QY 1308 TGAATATAAATGGTGGAAATGTCTCTTGGAACAATGAAAACTTCTCTGCTTCAATGTGG 1367  
Db 1320 TGAATACAAATGGTGGAAACAGCTGCCCAACAACATGGAACGACCCGTTCTCACTGGGAT 1379  
QY 1368 GAATTCAAAGTCCGATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427  
Db 1380 CAACTTCGAGTACAAGGCCATGACAGGCTGGGAGGTGGCTGGTGATCACATTTACACAGC 1439  
QY 1428 GGCTGGAGGTTCTGCACAAATGATTACCTGATCTTAACTTGCATATCCAGGATTTAAACC 1487  
Db 1440 TGCTGGAGCCTCAGACAAATGACTTCATGATTCTCACTCTGGTTGTGCCAGGATTTAGACC 1499  
QY 1488 ACCAACATCTATGACTGGAGCCAC--GGGTTCTGAACTAGGAAGAATAACATTTGTCTT 1544  
Db 1500 TCCGCAGTCGGTGATGGCAGACACAGAGAATAAAGAGGTGGCCAGAAATCACATTTGTCTT 1559  
QY 1545 TGAGACCCCTCTGTTCAGCTGACTGTGTTTTGTACTTTCATGGTGGATATTAAATAGAAAAAG 1604  
Db 1560 TGAGACCCCTCTGTTCTGTGAACCTGTGAGCTCTACTTTCATGTTGGTGGTGTAATCTTAGGAC 1619  
QY 1605 TACAAATGTGGTAGAATCGTGGGGTGGAAACCAGAAAAACAAGCTTACACCCATATCAT 1664  
Db 1620 CAACACTCCTGTGGAGACGTGGAAAGGTTCCAAAGGCAAAACAGTCCCTATACCTACATCAT 1679  
QY 1665 CTTCAAGAATGCAACTTTTACATTTACATGGGCATTTCCAGAGAACTAATCAGSGTCAAGA 1724  
Db 1680 TGAGGAGAACACTACCACGAGCTTCACCTGGGCCCTCCAGAGGACCACCTTTTCATGAGGC 1739  
QY 1725 TAATAGACGGTTTCATCAATGACATGGTGAAGATTATTCTATCAGAGCCCACTAATGCAGT 1784  
Db 1740 AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCACCAATGTTAT 1799  
QY 1785 TGATGGGGTGGCGTCCCTCATGCCGTGCCCTGTGCCCTAGAACCCCTCTGATGTGGGCTCCTC 1844  
Db 1800 GAATGGCGTGGCCTCCTACTGCGCTGCCCTGTGCCCTAGAACCCCTCTGATGTGGGCTCCTC 1859  
QY 1845 GTGTGTCCCTGCCCTCCAGGCCCACTACATTGAGAAAGAAACCAACCAGTGCAAGGAATG 1904  
Db 1860 CTGCACCTCTTGTCTCTGCTGGTTACTATATTGACCGAGATTCAGGAACCTGCCACTCCTG 1919  
QY 1905 TCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGATTTCCATG 1964  
Db 1920 CCCCCCTAACACAAATTCTGAAAGCCCCACCAGCCTTATGGTGTCCAGGCCCTGTGTGCCCTG 1979  
QY 1965 CGGGCCTGGGAGTAAAAACAATCAGGACCATTCCGGTTTGCTATAGTGACTGCTTTTCTA 2024

Db 1980 TGGTCCAGGGACCAAGAACAAAGATCCACTCTCTGTGCTACAATGATTGCACCTTCTC 2039  
QY 2025 CCATGAAAAAGAAAAATCAGATTTTGCACATATGACTTTTAGCAACCTCAGCAGTGTGGGCTC 2084  
Db 2040 ACGCAACACTCCAACCCAGGACTTTCAACTACAACTTCTCCGCTTTGGCAAAACACCGTCA 2099  
QY 2085 ATTAAATGAATGGCCCCAGCTTCACCTCCAAGGGAACAAAATACTTCCATTCTTCAATAT 2144  
Db 2100 TCTTGTCTGGAGGGCCAGCTTCACTTCCAAGGGTTGAAAATACTTCCATCACCTTACCCCT 2159  
QY 2145 CAGTTTATGTGGGCATGAGGGGAAGATGGCTCTCTGTACCAACAATATAAACAGACTT 2204  
Db 2160 CAGTCTCTGTGGAACCCAGCGGTAGGAAAATGTCTGTGTCACCGACAATGTCACTGACCT 2219  
QY 2205 TACAGTAAAAGAAAATAGTGGCAGGGTCAGATGATTACACAAAATTTGGTAGGGGCATTTGT 2264  
Db 2220 CCGGATTCTTGAGGGTGAGTCAAGG-----TTCTCCAATCTATCACAGCCTACGT 2270  
QY 2265 ATGCCAGTCAACAAATTTATTCCTTCTGAAAGTAAGGTTTCCGAGCAGCCTTTATCATCACA 2324  
Db 2271 CTGCCAGGCAGTCAATCTCCCCCAGAGGTTGACAGGCTACAAGGCCGGGTTTCTCTCACA 2330  
QY 2325 ATCCATCATTTCTGGCAGATACATTTCATAGGAGTCACAGTTGAAACCCACATTTGAAAAATAT 2384  
Db 2331 GCCTGTCAAGCCTTGTGATCGACTTATTGGGGTGCACAACAGATATGACTCTGGATGGANT 2390  
QY 2385 TAATATAAAAGAAGATATGTTCCAGTTCCAAACAGCCAAANTACCAGATGTGCATTTCTT 2444  
Db 2391 CACCTCCCAGCTGAACCTTTTCCACCTGGAGTCCCTTGGGAATACCGGACGTGATCTTCTT 2450  
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAATGGCCGATCAACTGCTGTGAAAAAT 2504  
Db 2451 TTATAGGTCCAATGATGTGACCCAGTCCCTGCAGTTCTGGGAGATCAACCCACCATCCGCGT 2510  
QY 2505 GAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCCCAGCAAGTCCCCCAGC 2564  
Db 2511 CAGGTGCAGTCCACAGAAAAACTGTCCCTGGAAAGTTTGCTGCTGCCAGGAACGTGCTCAGA 2570  
QY 2565 AGGTACCTGTGATGGGTGTACGTTCTATTTCCTGTGGGAGAGTGCTGAAGCTTGCCTCTT 2624  
Db 2571 TGGGACCTGTGATGGCTGCAACTTCCACTTCCCTGTGGGAGAGCGCGGCTGCTTGCCTGCT 2630  
QY 2625 GTGTACGGAGCATGACTTCCATGAGATTTGAGGGAGCCTGCAAGAGAGAGGATTTTCAGGAAC 2684  
Db 2631 CTGCTCAGTGGCTGACTTACCATGCTATCGTCAGCAGCTGTGTGGTGGGATCCAGAGAAG 2690  
QY 2685 CTTGTATGTGTGGAATGAACCTAAATGGTGCATTAAGGGAATTTCTTTGCCTGAGAAAAA 2744  
Db 2691 TACTTACGTGTGGCGAGAAACCCCAAGCTATGCTCTGGTGGCATTTCTCTGCCTGAGCAGAG 2750  
QY 2745 GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGGTGGAGCCGGTGTGGGAGCTTT 2804  
Db 2751 AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGCATCTCTGCAGGCACCTG 2810  
QY 2805 TACTGCCGTTTTCGTGGTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATCAAAAAAGAAAA 2864  
Db 2811 TACTGCCATCCTGCTCACCGTCTTGACCTGCTACTTTTGGAAAAAGAAATCAAAAACCTAGA 2870  
QY 2865 GAAGACCATTTTGAATCTGTTCA 2887  
Db 2871 GTACAAAGTACTCCAACCTGGTGA 2893

RESULT 8  
AAS21262  
ID AAS21262 standard; cDNA; 3501 BP.  
XX AAS21262;  
AC AAS21262;  
XX  
DT 24-OCT-2001 (first entry)  
XX  
DE Human cDNA sequence encoding for PRO4985 polypeptide.  
XX





Db 1158 CGATGCCAACGGAGAGACACAACACTCATGTACAAATGGGCCAAGCCGAAAAATCTGTAGCGA 1217

QY 1128 GGATCTCACAGATGCTATATAGATTGCCCCCTTCTGGAGAGAGAAGGATTGTCCGCCCTTG 1187

Db 1218 GGACCTTGAGGGGCGAGTGAAGCTGCCTGCCTCTGGTGAAGACCCACATGCCACCCCTG 1277

QY 1188 CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGGCATCCCTGTCTCCTCGGAACATT 1247

Db 1278 CAACCCAGGCTTCTTCAAAACCAACAACAGCACCTGCCAGCCCTGCCCATATGTTCTCTA 1337

QY 1248 TTCAGATGGAACCAAGAAATGTAGACCATGTCCAGAGGAACGGAGCCTGCACTTGGCTT 1307

Db 1338 CTCCAATGG--CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGATT 1394

QY 1308 TGAATATAAATGGTGGAAATGTCTCTTCTGGCAACATGAAAACTTCTCTGCTTCAATGTTGG 1367

Db 1395 TGAATACAAATGGTGGAAACACGCTGCCCCACAACATGGAACGACCGTTCTCAGTGGGAT 1454

QY 1368 GAATTCAAAGTGCATGGAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427

Db 1455 CAACTTCGAGTACAAGGGCATGACAGGCTGGGAGGTGGCTGGTGATCACATTTACACAGC 1514

QY 1428 GGCTGGAGGTTCTGACATGATTAACCTGATCTTAACTTGCATATCCCAGGATTTAAACC 1487

Db 1515 TGCTGGAGCCTCAGACAATGACTTCATGATTTCTCACTCTGTTGTGCCAGGATTTAGACC 1574

QY 1488 ACCAACATCTATGACTGGAGCCAC--GGGTTCTGAACTAGGAAGAATAACATTTGTCTT 1544

Db 1575 TCCGAGTCGGTGATGGCAGACACAGAGAATAAAGAGGTGGCCAGAAATCACATTTGTCTT 1634

QY 1545 TGAGACCCCTCTGTTTCAGCTGACTGTGTTTTGTACTTTCATGGTGGATATTAAATAGAAAAAG 1604

Db 1635 TGAGACCCCTCTGTTCTGTGAATGTGAGCTCTACTTCTATGGTGGTGTGAATTTCTAGGAC 1694

QY 1605 TACAAATGTGTAGAAATCGTGGGTGGAACCAAGAAAGAAAAACAAGCTTACACCCCATATCAT 1664

Db 1695 CAACACTCCTGTGGAGACGTGGAAGGTTCCAAAGGCAACAGATCCTATACCTACATCAT 1754

QY 1665 CTTCAGAAATGCAACTTTTACATTTACATGGGCATTCCAGAGAACTAATCAGGGTCAAGA 1724

Db 1755 TGAGGAGAACACTACCACGAGCTTCACCTGGGCCCTCCAGAGGACCACTTTTTCATGAGGC 1814

QY 1725 TAATAGACGGTTTCATCAATGACATGGTGAAGATTTATTTATATCACAGCCACTAATGCAGT 1784

Db 1815 AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCACCAATGTTAT 1874

QY 1785 TGATGGGGTGGCGTCTCATGCGCGTGCCTGTGCCCCCTGTGAAAGCCCTCTGATGTGGGCTCCTC 1844

Db 1875 GAATGGCGTGGCCTCCTACTGCGCGTCCCTGTGCCCCCTAGAACCCCTCTGATGTGGGCTCCTC 1934

QY 1845 GTGTGTCCCCCTGCCCTCCAGGCCACTACATTTAGAAAGAAACCAACCCAGTGCAGGAATG 1904

Db 1935 CTGCACCTCTGTCTGCTGTTACTATATTGACCGAGATTCAGGAACCTGCCACTCCTG 1994

QY 1905 TCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTTCCATG 1964

Db 1995 CCCCCCTAACACAATTTCTGAAAGCCCCACAGCCCTTATGGTGTCCAGGCCCTGTGTGCCCTG 2054

QY 1965 CGGGCCTGGGAGTAAAAACAATCAGGACCATTTCGGTTTGTCTATAGTACTGCTTTTCTA 2024

Db 2055 TGGTCCAGGGACCAAGAACAACAGATCCACTCTCTGTGTACAAATGATTGCACCTTCTC 2114

QY 2025 CCATGAAAAAGAAATCAGATTTTGCACATGACTTTAGCAACCTCAGCAGTGTGGGCTC 2084

Db 2115 ACGCACACTCCAACCCAGGACTTTCAACTACAACCTTCTCCGCTTTGGCAAAACACCCGTCAC 2174

QY 2085 ATTAATGAATGGCCCCAGCTTCACCTCCAAAGGAACAAATACTTCCATTTCTTCAATAT 2144

Db 2175 TCTTGTGGAGGGCCCAAGCTTCACTTCCAAAGGGTTGAAATFACCTTCCATCACTTTACCCCT 2234

QY 2145 CAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATAATACAGACTT 2204

Db 2235 CAGTCTCTGTGGAAACCCAGGGTAGGAAAAATGTCTGTGTGACCGGACAAATGTCACTGACCT 2294

QY 2205 TACAGTAAAAAGAAATAGTGGCAGGGTCAGATGATTAACAAAAATTTGGTAGGGCAATTTGT 2264

Db 2295 CCGGATTCTTGAGGGTGAGTCAGGG-----TTCTCCAAATCTATCACAGCCTACGT 2345

QY 2265 ATGCCAGTCAACAATTAATTCCTTCTGAAAGTAAGGTTTCCGAGCAGCCTTAATCATCACA 2324

Db 2346 CTGCCAGGCACTCATCATCCCCCAGAGGTGACAGGCTACAAGGCCGGGTTTCTCTCACA 2405

QY 2325 ATCCATCATTTCTGGCAGATACATTCATAGGAGTCACAGTTGAAACCACATTTGAAAAATAT 2384

Db 2406 GCCTGTCAGCCTTGTGATCGACTTATTGGGGTGACAACACAGATATGACTCTGGATGGAAT 2465

QY 2385 TAATATAAAAGAAGATATGTTCCAGTTTCCAACAAGCCAAATACCAGATGTGCATTTCTTT 2444

Db 2466 CACCTCCCAGCTGAACCTTTTCCACCTGGAGTCCCTGGGAATACCGGACGTGATCTTCTT 2525

QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAATGGCCCGATCAACTGCTGTGAAAAAT 2504

Db 2526 TTATAGGTCCAATGATGTGACCCAGTCTTCAGTTCTGGGAGATCAACCACCAATCCGCGT 2585

QY 2505 GAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCCAGCAAGTGCCACAGC 2564

Db 2586 CAGGTGCAGTCCACAGAAAACTGTCCCTGGAAGTTTGCCTGCCAGGAACGTGCTCAGA 2645

QY 2565 AGGTACCTGTGATGGGTCTACGTTCTATTTCCCTGTGGAGAGTGTGAAGCTTGCCCTCT 2624

Db 2646 TGGGACCTGTGATGGCTGCAACTTCCACTTCCCTGTGGAGAGCGCGGCTGCTTGCCCGCT 2705

QY 2625 GTCTACGGAGCATGACITTCATGAGATTGAGGGAGCCTGCCAAGAGAGGATTTTCAGGAAAC 2684

Db 2706 CTGCTCAGTGGCTGACTACCATGCTATCGTCAGCAGCTGTGTGGCTGGGATCCAGANGAC 2765

QY 2685 CTGTATGTGTGGAATGAACTTAATGGTGCATTTAAGGAATTTCTTTGCCCTGAGAAAAA 2744

Db 2766 TACTTACGTGTGNCGAGAACCCAAAGCTATGCTCTGTGGTGCATTTCTCTGCCCTGAGCAGAG 2825

QY 2745 GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGTGGGAGCCGGTGTGGGAGCTTT 2804

Db 2826 AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCGACCTG 2885

QY 2805 TACTGCCCGTTTGTGGTGGTCTGACCTGTGACCTGTACTTCTGGAAAAAGAAATCAAAAAA 2864

Db 2886 TACTGCCCATCTGTCTACCCGTCTTGACCTGTACTTTTGGAAAAAGAAATCAAAAACTAGA 2945

QY 2865 GAAGACCATTTTGAATCTGTCTCA 2887

Db 2946 GTACAAAGTACTCCAAGCTGTGTA 2968

RESULT 9

AAF89775

ID AAF89775 standard; DNA; 2610 BP.

XX

AC AAF89775;

XX

DT 23-JUL-2001 (first entry)

XX

DE Nucleotide sequence of a human protein expressed in tumour cells.

XX

KW Tumour cell; immunological disease; autoimmune disease; cancer;

KW infection; ss.

XX

OS Homo sapiens.

XX

FH Location/Qualifiers

FT CDS

FT /tag= a

FT /product= "tumour expressed protein"

XX

PN WO200131003-A1.

XX

PD 03-MAY-2001.







QY 411 TGAATGGGATGAATTGCCGGCAGGATTTCTAAACATCGCAACATTCATGGACACTGTGGT 470  
Db 375 TGAGTGGGATGAGCTGCCCATGGCTTTGCCAGCCTCTCAGCCAACATGGAGCTGGATGA 434  
QY 471 GGGCCCTCTGCACAGCAGGCCGAGCTGTAACAACTCTTCTGGATCCCTCGTGGAAA 530  
Db 435 CAGTGTCTGAGTCCA--CCGGGAAGTGTACTTCGTCCCAAGTGGGTTCGCCGGGGCGA 491  
QY 531 CTACATAGAAATCTAATCGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAA 590  
Db 492 CTACATCGCTCCAACACGGACGGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 551  
QY 591 GAAGTCAGGCTATGTCTTCTTTGAGTACCAGTATGTCGACAAACACATCTTCTTTGAGTT 650  
Db 552 GCAATCTGGCACCCTTAACCTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 611  
QY 651 CTTTATTCAAATGATCAGTGCCAGGAGATGGACACACCACCTGACAAAGTGGGTAAAACT 710  
Db 612 TTTCGTTCAGAAATGACCAGTGCCAGGCCCAATGCAGATGACTC---CAGGTGGATGAAGAC 668  
QY 711 TACAGACAATGGAGAATGGGGCTCTCATTTCTGTAATGCTGAATFCAGGCACAAACATACT 770  
Db 669 CACAGAGAA---AGGATGGGAATTCACAGAGTGTGGAGCTAAATCAGGGCAATAATGTCTT 725  
QY 771 CTACTGGAGAACTACAGGCATCCTTATGGGTCTTAAGCGGTCTAAGCTGTGCTGGTAAA 830  
Db 726 CTATTGGAGAAACACAGCCTTCTCAGTATGGACCAAGTACCCAAAGCTTCCCTTGCAAACCTGGCAC 845  
QY 831 AAATATCACAAATTGAAGGGGTGGCGTACACATCAGAAATGTTTTCTTGCAGGCCAGGCAC 890  
Db 786 AAACATTGCCATTAACAGGGTGGCCTACACTTCAGAAATGCTTCCCTTGCAAACCTGGCAC 845  
QY 891 ATTACGCAACAAACACAGGTTTCATTCAACTGCCAGGTGTGTCGCCAGAAACACCTATTCTGA 950  
Db 846 GTATGCAGACAAAGCAGGGCTCCTCTTTCTGCAAACTTTGCCAGGCCAACTCTTATTCAA 905  
QY 951 GAAAGGAGCCAAAGAAATGTATAAGGTGTA--AGACGACTCTCAATTTTCAGGATCCAG 1007  
Db 906 TAAAGGAGAAACTTCTTGCCACCAGTGTGACCTTGACAAATACTCAGAGAAAGGATCTTC 965  
QY 1008 TGAGTGTACAGAGCCCTCCTGTATCCACAAAGACTATTTCCAGATCCATACTCCATG 1067  
Db 966 TTCTGTAACTGCGCCACAGCTTGACAGACAAAGATTTATTTCTACACACACACGCGCTG 1025  
QY 1068 TGATGAAGAAAGGAAGACACAGATAATGTACAAAGTGGATAGAGCCCAAAATCTGCCGGGA 1127  
Db 1026 CGATGCCAACGGAGACACAACTCATGTACAAATGGCCAAAGCCGAAATCTGTAGCGA 1085  
QY 1128 GGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAAGAAGGATTTGCCGCTTG 1187  
Db 1086 GGACCTTGAGGGGGCAGTGAAGCTGCCCTGCCTGTGGTGAAGACCCACTGCCACCCCTG 1145  
QY 1188 CAACCCCTGGATTTATAACAAATGGATCATCTTCTTGCCATCCCTGTCTCCTCGGAACATT 1247  
Db 1146 CAACCCAGGCTTCTCAAAACCAACACAGCACCTGCCAGCCCTGCCCATATGGTCCCTA 1205  
QY 1248 TTCAGATGGAACCAAGAAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTT 1307  
Db 1206 CTCCAATGG--CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGTGTGGGATT 1262  
QY 1308 TGAATATAAATGGTGAATGTCTTCTGCGCAACATGAAACCTTCCGTCTCAATGTTGG 1367  
Db 1263 TGAATACAAATGGTGAACACGCTGCCACAAACATGGAACGACCCGTTCTCAGTGGGAT 1322  
QY 1368 GAATTCAAAGTCGATGGAATGAATGGTTGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427  
Db 1323 CAACTTCGAGTACAAAGGCATGACAGGCTGGGAGGTGGCTGGTGATCATATTTACACAGC 1382  
QY 1428 GGCTGGAGGTTCTGACAATGATTACCTGATCTTAAACTTGCAATATCCCAGGATTTAAACC 1487  
Db 1383 TGCTGGAGCCTCAGACAATGACTTTCATGATTTCTACTCTGTTGTCGCCAGGATTTAGACC 1442  
QY 1488 ACCAACATCTATGACTGGAGCCAC--GGGTTCTGAACTAGGAAGAATAACATTTGTCTT 1544

Db 1443 TCCGCAGTCGGTGATGGCAGACACAGAGAATAAAGAGGTTGGCCAGAATCACATTTGTCTT 1502  
QY 1545 TGAGACCCCTCTGTTCAGCTGACTGTGTTTTGTACTTTCATGGTGATATTAATAGAAAAAG 1604  
Db 1503 TGAGACCCCTCTGTTCTGTGAACTGTGAGCTCTACTTTCATGGTGGTGTGAATCTAGGAC 1562  
QY 1605 TACAAATGTGGTAGAATCGTGGGTGGAACCAAGAAAGAAACAAAGCTTACACCCATATCAT 1664  
Db 1563 CAACACTCTGTGGAGACGTGGAAGGTTCCAAAGGCAACAGTCCCTATACCTACATCAT 1622  
QY 1665 CTTCAAGAATGCAACTTTTACATTTACATGGGCATTCAGAGAACTAATCAGGGTCAAGA 1724  
Db 1623 TGAGGAGAACACTACCACGAGCTTCACCTGGGCTTCCAGAGGACCACCTTTTCATGAGGC 1682  
QY 1725 TAAATAGACGGTTTCATCAATGACATGGTGAAGATTTATTTCTATCACAGCCACTAATGCAGT 1784  
Db 1683 AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCACCAATGTTAT 1742  
QY 1785 TGATGGGTGGCGTCCCTCATGCCGTGCCCTGTGCCCTCGTTCTGAAACAGTCGGGTTTCATC 1844  
Db 1743 GAATGGCGTGGCCTCCTACTGCCGTGCCCTGTGCCCTTAGAAGCCTCTGATGTGGGCTCCTC 1802  
QY 1845 GTGTGTCCCTGCCCTCCAGGCCACTACATTTGAGAAAGAAACCAACACAGTGCAAGGAATG 1904  
Db 1803 CTGCACCTCTTGTCCTGCTGTTACTATATTGACCGAGATTCAGGAACCTGCCACTCCTG 1862  
QY 1905 TCCACCTGACACCTACCTGTCTCATATACATCAGGTCTATGGCAAAGAGGGCTTGTATTCATG 1964  
Db 1863 CCCCCCTAACACAAATTTGAAAGCCCAACAGCCTTATGGTGTCCAGGCCCTGTGTGCCCTG 1922  
QY 1965 CGGGCCTGGGAGTAAAAACAATCAGAGCCATTCGGTTTGTATAGTGACTGCTTTTCTA 2024  
Db 1923 TGGTCCAGGACCAAGAAACAACAAGATCCACTCTCTGTGTACAAATGATTGCACCTTCTC 1982  
QY 2025 CCATGAAAAAGAAAAATCAGATTTTGCACTATGACTTTAGCAACCTCAGCAGTGTGGGCTC 2084  
Db 1983 ACGCAACACTCCAACCCAGGACTTTCAACTACAACTTCTCCGCTTTGGCAACACCCGTCAC 2042  
QY 2085 ATTAATGAATGGCCCCCAGCTTCACCTCCAAAGGAACAATAATCTTCCATTTCTTCAATAT 2144  
Db 2043 TCTTGTGTGGAGGCCAAGCTTCACCTTCCAAAGGGTTGAAATACTTCCATCATTACCTT 2102  
QY 2145 CAGTTTATGTGGGCATGAGGGGAAGAGATGGCTCTCTGTACCAACAATATAACAGACTT 2204  
Db 2103 CAGTCTCTGTGGAAACCAGGGTAGGAAAATGTCTGTGTGCACCCGACAATGTCACTGACCT 2162  
QY 2205 TACAGTAAAGAAATAGTGGCAGGTCAGATGATTACACAAAATTTGGTAGGGGCATTTGT 2264  
Db 2163 CCGGATTCCTGAGGGTGAGTCAGGGTT-----CTCCAAATCTATCACAGCCCTACGT 2213  
QY 2265 ATGCCAGTCAACAATTTATTCCTTCTGAAAGTAAGGGTTTCCGAGCAGCCTTATCATCACA 2324  
Db 2214 CTGCCAGGCAGTCATCATCCCCCAGAGGTGACAGGCTACAAGGCCGGGGTTTCTCTCACA 2273  
QY 2325 ATCCATCATCTCTGGCAGATACATTCATAGGAGTCACAGTTGAAACACACATTTGAAAAATAT 2384  
Db 2274 GCCTGTACGCTTGTGTATCGACTTATTGGGGTGACAACAGATATGACTCTGGATGGAAT 2333  
QY 2385 TAATATAAAGAGATATGTTCCAGTTTCCAAACAAAGCCAAATACCAAGATGTGATTTCTT 2444  
Db 2334 CACCTCCCCCAGCTGAACCTTTTCCACCTGGAGTCCCTTGGGAATACCCGACGTGATCTTCTT 2393  
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAAATGGCCCGATCAACTGCTGTGAAAAAT 2504  
Db 2394 TTATAGGTCCCAATGATGTGACCCAGTCCCTGCAGTTCTGGGAGATCAACACCATCCGCGT 2453  
QY 2505 GAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTTCACTCCCAGCAAGTGCCCCAGC 2564  
Db 2454 CAGGTGCAGTCCACAGAAAAACTGTCCCTGGAAGTTTGTGCTGCCAGGAACGTGCTCAGA 2513  
QY 2565 AGGTACCTGTGATGGGTGTACGTTTCTATTCTCTGTGGGAGAGTGCTGAAGCTTGCCCTCT 2624

Db	2514	TGGGACCTGTGATGGCTGCAACTTCCACTTCCTGTGGGAGAGCGCGGCTGCTTGCCCGCT	2573
QY	2625	GTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCTGCAAGAGAGAGGATTTTCAGGAAAC	2684
Db	2574	CTGCTCAGTGGCTGACTACCGTGCTATCGTCAGCAGCTGTGTGGCTGGGATCCAGAAGAC	2633
QY	2685	CTTGCTATGTGTGGAATGAACCTAAATGGTGCATTAAAGGAATTTCTTTGCCCTGAGAAAAA	2744
Db	2634	TACTTACGTGTGGCGAGAACCCCAAGCTATGCTCTGGTGGCATTTCTCTGCCTGAGCAGAG	2693
QY	2745	GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGGTG	2784
Db	2694	AGTCACCATCTGCAAAAACCATAGATTTCTGGCTGAAAGTG	2733
RESULT 11			
AAF27997			
ID	AAF27997 standard; DNA; 2554 BP.		
XX	AC AAF27997;		
XX	08-MAY-2001 (first entry)		
DT	Human TR13 receptor coding sequence SEQ ID NO: 1.		
XX	Human; tumour necrosis factor receptor; TR13; TR14; infection;		
KW	cancer; autoimmune disease; allergy; inflammatory disease;		
KW	graft rejection; apoptosis; cardiovascular disease; aneurysm; ds.		
XX	Homo sapiens.		
OS	WO200105834-A1.		
XX	25-JAN-2001.		
PD	14-JUL-2000; 2000WO-US19343.		
XX	16-JUL-1999; 99US-0144087.		
PR	18-AUG-1999; 99US-0149450.		
PR	20-AUG-1999; 99US-0149712.		
PR	10-SEP-1999; 99US-0153089.		
XX	(HUMA-) HUMAN GENOME SCI INC.		
PA	Ruben SM, Ni J, Young PE;		
XX	WPI; 2001-112682/12.		
DR	P-PSDB; AAB35328.		
XX	Nucleic acids encoding 2 human tumor necrosis factor receptor		
PT	polypeptides ((TR13) and (TR14)), useful for the prevention, diagnosis		
PT	and treatment of, e.g. cancers, acquired immune deficiency syndrome and		
PT	hypohidrotic ectodermal dysplasia -		
XX	Claim 2; Page 366-369; 418pp; English.		
XX	The present invention provides the protein and coding sequences of the		
CC	human tumour necrosis factor receptors TR13 and TR14. These sequences are		
CC	useful in the diagnosis and treatment of many diseases, including cancer,		
CC	autoimmune diseases, cardiovascular disorders, allergies,		
CC	neurodegenerative diseases, graft rejection, inflammation, aneurysms and		
CC	infections.		
XX	Sequence 2554 BP; 642 A; 722 C; 588 G; 602 T; 0 other;		
SQ			
Query Match 20.3%; Score 686.6; DB 22; Length 2554;			
Best Local Similarity 59.6%; Pred. No. 1e-169;			
Matches 1256; Conservative 0; Mismatches 834; Indels 19; Gaps 5;			
QY	785	CAGGCATCCTTATGGGTTCTAAGCGGTCAGCCCTGTGCTGGTAAAAAATATCACAATTG	844
Db	18	CAGCCTTCTCAGTATGGACCAAAAGTACCCAAGCCTGTGCTGGTGAGAAACATTGCCATAA	77

QY	845	AAGGGTGGCGGTACACATCAGAAATGTTTCTTGTGCAAGCCAGGCACACATTTCAGCAACAAC	904
Db	78	CAGGGTGGCCTACACTTCAGAAATGCTTCCCCTGCAAAACCTGGCACGTATGCAGACAAGC	137
QY	905	CAGGTTCAATCAAACTGCCAGTGTGTCTCCAGAAACACCTATCTTGAGAAAGGAGCCAAAG	964
Db	138	AGGGCTCCTCTTCTGCAAACTTG-CCCAGCAACTCTTATCAAAATAAAGGAGAAACTT	196
QY	965	AATGTATAAGGTGTAA--AGACGACTCTCAATTTTCAGGATCCAGTGAGTGTACAGAGC	1021
Db	197	CTTGCCACCAGTGTGACCCCTGCACAAATACTCAGAGAAAGGATCTTCTTCTGTAACTGC	256
QY	1022	GCCCTCCTGTACCACAAAAGACTATTTCCAGATCCATCTCCATGTGATGAAGAAAGSAA	1081
Db	257	GCCCAGCTTGACACAGACAAAGATTATTCTACACACACACGCGCTCGGATGCCAACGAG	316
QY	1082	AGACACAGATAATGTACAAGTGGATAGAGCCCCAAAATCTGCCGGGAGGATCTCACAGATG	1141
Db	317	AGACACAACCTCATGTACAAATGGGCCAAGCCGAAAATCTGTAGCGAGGACCTTGAGGGGG	376
QY	1142	CTATTAGATTGCCCCCTTCTGGAGAGAAGAGGATTGTCCGCTTGCAACCCCTGGATT	1201
Db	377	CAGTGAAGCTGCCCTGCCCTGTGGTGTGAAGACCCACTGCCACCCCTGCAACCCAGGCT	436
QY	1202	ATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTCTGGAACATTTTCAGATGGAACA	1261
Db	437	TCAAAACCAACAACAGCACCTGGCAGCCCTGCCCATATGGTTCCTACTCCAAATGG--CT	493
QY	1262	AAGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTGAATATAAATGGT	1321
Db	494	CAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGATTGAATACAAATGGT	553
QY	1322	GGATGTCCTTCTTGCAACATGAAAAACTTCTCTGCTTCAATGTGGGAATTCAAAGTGGC	1381
Db	554	GGAACACGCTGCCACAAACATGAAAACGACCGTTCTCAGTGGGATCAACTTCGAGTACA	613
QY	1382	ATGGAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTC	1441
Db	614	AGGCATGACAGGCTGGGAGGTGGTGGTATCACATTTACACAGCTGCTGGAGCCCTCAG	673
QY	1442	ACAATGATTACCTGATCTTAACTTGCATATCCCAGGATTTAAACCCACCACATCTATGA	1501
Db	674	ACAATGACTTCATGATTCCTCACTCTGGTGTGCCAGGATTTAGACCTCCGCAGTCGGTGA	733
QY	1502	CTGGAGCCAC--GGGTTCTGAACTAGGAAGAATAACATTTGCTTTTGAGACCCCTCTGT	1558
Db	734	TGGCAGACACAGAGAATAAAGAGGTGGCCAGAAATCACATTTGCTTTGAGACCCCTCTGT	793
QY	1559	CAGCTGACTGTGTTTGTACTTTCATGGTGGATATTAATAGAAAAGTACAAATGTGTAG	1618
Db	794	CTGTAAACTGTGAGCTCTACTTTCATGGTGGTGTGAATTTCTAGGACCAACACTCCTGTGG	853
QY	1619	AATCGTGGGTGGAAACCAAGAAAACAAAGCTTACACCCCATATCATCTTCAAGAAATGCAA	1678
Db	854	AGACGTGGAAAGGTTCCAAAAGGCAAAACAGTCCCTATACCTACATCATTTGAGGAGAACA	913
QY	1679	CTTTTACATTTACATGGGCATTCACAGAGAACTAATCAGGGTCAAGATAATAGACGGTTCA	1738
Db	914	CCACGAGCTTCACCTGGGCTTCCAGAGGACCACCTTTTCATGAGGCAAGCAGGAAGTACA	973
QY	1739	TCAATGACATGGTGAAGATTTATTTCTATCACAGCCACTAATGCAGTTGATGGGGTGGCT	1798
Db	974	CCAATGACGTTGCCAAGATCTACTCCATCAATGTACCACCAATGTATGAATGGCGTGGCT	1033
QY	1799	CCTCATGCCGTGCCCTGTGCCCTTCTGAACAGTCCGGTTCATCGTGTGTCCTCCCTGCC	1858
Db	1034	CCTACTGCCGTCCCCTGTGCCCTAGAAGCCCTCTGATGTGGGCTCCTCCTGCACCTCTTGT	1093
QY	1859	CTCCAGGCCACTACATTGAGAAAGAAACCAACCAGTGCAGGAATGTCCACCTGCACACCT	1918
Db	1094	CTGCTGGTTACTATATTGACCGAGATTCAGGAACCTGCCACTCCTGCCCTTACACAA	1153
QY	1919	ACCTGTCCATACATCAGGTCTATTGGCAAGAGGGCTTGTATTCCATGCGGGGCTGGGAGTA	1978

Db 1154 TTCTGAAAGCCACCAGCCCTATGGTGTCAGGCCCTGTGTGCCCTGTGGTCCAGGGACCA 1213  
QY 1979 AAAACAATCAGGACCATTCGGTTTGCTATAGTGACTGCTTTTCTACCATGAAGAAAGAAA 2038  
Db 1214 AGAACACAAGATCCACTCTCTGTGCTACAATGATTGCACCTTCTCAGCGAACACTCCAA 1273  
QY 2039 ATCAGATTTTGCACTATGACTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGGCC 2098  
Db 1274 CCAGGACTTCAACTACAACTTCTCCGCTTTGGCAACACCGTCACTCTTGCTGGAGGGC 1333  
QY 2099 CCAGCTTCACCTCCAAAGGAACAAAATACTTCCATTCTTCAATATCAGTTTATGTGGGC 2158  
Db 1334 CAAGCTTCACCTCCAAAGGGTTGAAATACTTCCATCACTTACCCCTCAGTCTCTGTGGAA 1393  
QY 2159 ATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTTTACAGTAAAGAAA 2218  
Db 1394 ACCAGGTAGGAAAATGTCTGTGTCACCGACAATGTCACTGACCTCCGGATTCTCCTGAGG 1453  
QY 2219 TAGTGGCAGGGTCAGATGATTACACAAAATTTGGTAGGGGCAATTTGTATGCCAGTCAACAA 2278  
Db 1454 GTGAGTCAGGG-----TTCTCCAAATCTATCACAGCCTACGTCTGCCAGGCAGTCA 1504  
QY 2279 TTATTCCCTTCTGAAAGTAAGGGTTTCCGAGCAGCCCTTATCATCACAAATCCATCATCTTCTGG 2338  
Db 1505 TCATCCCCCAGAGGTGACAGGCTACAAGCCCGGGTTTCTCTCACAGCCTGTGAGCCCTTG 1564  
QY 2339 CAGATACATTCATAGGAGTCACAGTTGAAACCACATTTGAATAATATTAAATATAAAGAA 2398  
Db 1565 CTGATCGACTTATTGGGGTGACACAGATATGACTCTGGATGGAATCACCTCCCCAGCTG 1624  
QY 2399 ATATGTTCCCACTTCCAAACAAGCAAAATACCAGATGTGCAATTTCTTTTATAAGTCTTCTA 2458  
Db 1625 AACTTTTCCACCTGGAGTCCCTTGGGAATACCGAGCTGATCTCTTTTATAGGTCCAATG 1684  
QY 2459 CAGCAACAACATCTTGTATTAAATGGCCGATCAACTGCTGTGAAAATGAGGTGTAATCCTA 2518  
Db 1685 ATGTGACCCAGTCTCTGCAGTTCTGCGAGATCAACCACCATCCGCGTCAGGTGCAGTCCAC 1744  
QY 2519 CTAATCTGGAGCAGGAGTGATTTTCACTCCCCAGCAAGTGCCCAAGCAGGTACCTGTGATG 2578  
Db 1745 AGAAAACTGTCCCTGGAAGTTTGTCTGCCAGGAACGTGCTCAGATGGGACCTGTGATG 1804  
QY 2579 GGTGTACGTTCTATTTCCTCTGGGAGAGTGCTGAAGCTTGGCCCTCTGTGTACGGAGCATG 2638  
Db 1805 GCTGCAACTTCCACTTCTCTGTGGGAGAGCGCGGCTGTGCCCCGCTCTGCTCAGTGGCTG 1864  
QY 2639 ACTTCCATGAGATTGAGGGAGCCTGCAAGAGAGGATTTTCAGGAACCTTTGTATGTGTGA 2698  
Db 1865 ACTACCATGCTATCGTCAGCAGCTGTGTGGCTGGGATCCAGAGACTACTTACGTGTGGC 1924  
QY 2699 ATGAACCTTAAATGGTGCAATTAAGGAATTTCTTTGCCCTGAGAAAAAGTTGGCAACCTGTG 2758  
Db 1925 GAGAACCCCAAGCTATGCTCTGTGGGCAATTTCTCTGCCCTGAGCAGAGAGTCAACCATCTGCA 1984  
QY 2759 AAACGGTTGACTTTTGGCTGAAGGTGGGAGCCGGTGTGGGAGCTTTTACTGCCGTTTTC 2818  
Db 1985 AAACCATAGATTCTGTGCTGAAAGTGGGCATCTCTGCAGGCACCTGTACTGCCATCCTGC 2044  
QY 2819 TGGTGGCTCTGACCTGTCTACTTCTGGAAGAAAGAAATCAAAAGAAAAAGAACACCATTTTGA 2878  
Db 2045 TCACCGTCTTGACCTGTCTACTTTTGGAAAAAGAAATCAAAACACTAGAGTACAAGTACTCCA 2104  
QY 2879 ATCTGTTCA 2887  
Db 2105 AGCTGGTGA 2113

RESULT 12  
AAX27340  
ID AAX27340 standard; DNA; 1129 BP.  
XX  
AC AAX27340;

XX 11-JUN-1999 (first entry)  
DT Human secreted protein gene 30 clone HODAZ26.  
XX  
DE Human; secreted protein; fusion protein; gene therapy; protein therapy;  
XX diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
XX  
OS Homo sapiens.  
XX  
PN WO9902546-A1.  
XX  
PD 21-JAN-1999.  
XX  
PF 07-JUL-1998; 98WO-US13684.  
XX  
PR 12-SEP-1997; 97US-0058785.  
PR 08-JUL-1997; 97US-0051916.  
PR 08-JUL-1997; 97US-0051918.  
PR 08-JUL-1997; 97US-0051919.  
PR 08-JUL-1997; 97US-0051920.  
PR 08-JUL-1997; 97US-0051925.  
PR 08-JUL-1997; 97US-0051926.  
PR 08-JUL-1997; 97US-0051928.  
PR 08-JUL-1997; 97US-0051929.  
PR 08-JUL-1997; 97US-0051930.  
PR 08-JUL-1997; 97US-0051931.  
PR 08-JUL-1997; 97US-0051932.  
PR 08-JUL-1997; 97US-0052732.  
PR 08-JUL-1997; 97US-0052733.  
PR 08-JUL-1997; 97US-0052793.  
PR 08-JUL-1997; 97US-0052795.  
PR 08-JUL-1997; 97US-0052803.  
PR 18-AUG-1997; 97US-0055684.  
PR 18-AUG-1997; 97US-0055722.  
PR 18-AUG-1997; 97US-0055723.  
PR 18-AUG-1997; 97US-0055947.  
PR 18-AUG-1997; 97US-0055948.  
PR 18-AUG-1997; 97US-0055949.  
PR 18-AUG-1997; 97US-0055950.  
PR 18-AUG-1997; 97US-0055953.  
PR 18-AUG-1997; 97US-0055954.  
PR 18-AUG-1997; 97US-0055964.  
PR 18-AUG-1997; 97US-0055984.  
PR 18-AUG-1997; 97US-0056360.  
PR 12-SEP-1997; 97US-0058660.  
PR 12-SEP-1997; 97US-0058661.  
PR 12-SEP-1997; 97US-0058664.  
XX  
(HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
PI Brewer LA, Ebner R, Fischer CL, Kyaw H, Lafleur DW, Li Y, Moore PA;  
PI Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR, Zeng Z;  
XX  
DR WPI: 1999-120770/10.  
DR P-PSDB; AAY02679.  
XX  
PT New isolated human genes and the secreted polypeptides they encode -  
PT useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders  
XX  
PS Claim 1; Page 262; 464pp; English.  
XX  
CC This sequence represents a nucleic acid molecule which encodes a  
CC secreted human protein. The gene number, and the clone it is derived  
CC from, are detailed in the descriptor line. The gene can be used to  
CC generate fusion proteins by linking to the gene to a human immunoglobulin





PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235834.  
PR 29-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

XX

PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI; 2001-488782/53.  
DR P-PSDB; AAU21345.

XX New polynucleotides and polypeptides for diagnosing, treating,  
PT preventing or prognosing e.g. diseases or disorders of the nervous,  
PT musculoskeletal, excretory, gastrointestinal, reproductive, and  
PT respiratory systems

PS Claim 1; SEQ ID No 689; 642pp; English.

XX  
CC The invention relates to novel nucleic acids encoding novel human foetal  
CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.  
CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. The antibodies to the antigens can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. Numerous  
CC examples of diseases and disorders treated by the nucleic acids and  
CC proteins are given in the specification. The present sequence

Query Match 14.5%; Score 492.6; DB 22; Length 672;  
Best Local Similarity 97.9%; Pred. No. 4.9e-119;  
Matches 514; Conservative 5; Mismatches 4; Indels 2; Gaps 2;

QY 2850 GAATCAAAAGAAAAGAACACCATTTTGAATCTGTTCAACTGAAACCTCAAGATCCCCA 2909

Db 150 GCAACCMAGGAAAAGRAGACCATTTTGAATCTGTTCAACTGAAAMCCTCAAGATCCCCA 209

QY 2910 AATATATGAAGACACAGTGTGTAGCCTTGAGACTAATGAACAAGAAACCTGCTCTAGT 2969

Db 210 AATATATGAAGACACAG-KCTGTAGCCTTGAGACTAATGAACAAGAAACCTGCTCTAGT 268

QY 2970 TTTACAGGACCATATTTAGGGTCTGTCTCATACCTGTACATTTGGTATGCTCACAGAG 3029

Db 269 TTTACAGGACCATATTTAGGGTCTGTCTCATACCTG-CACATTTGGTATGCTCACAGAG 327

QY 3030 GAGGGCCATGCCGCTGAAAAGGAGAGATTGAAACATTTGATTGCCTTATCACATGGT 3089

Db 328 GAGGGCCATGCCGCTGAAAAGGAGAGATTGAAACATTTGATTGCCTTATCACATGGT 387

QY 3090 CAAGTACCTTGCCAAATAAAGGAAAGCAATGATTTGGGTCTCAACTGAAGATGAAGCTC 3149

Db 388 CAAGTACCTTGCCAAATAAAGGAAAGCAATGATTTGGGTCTCAACTGAAGATGAAGCTC 447

QY 3150 AACTCAGGAAGAGATTTATCTGTATATACACATAACTGAAACCAAGTTTAAGCCCCACCA 3209

Db 448 AACTCAGGAAGAGATTTATCTGTATATACACATAACTGAAACCAAGTTTAAGCCCCACCA 507

QY 3210 ATGCACTGCTGATGCATGCCATATAATTAATGGGTAACCTTTTATCTTTATGATGTCTAC 3269

Db 508 ATGCACTGCTGATGCATGCCATATAATTAATGGGTAACCTTTTATCTTTATGATGTCTAC 567

QY 3270 ATAACAAGTGTGATTTGGAAGGCACATGTGACCATATGCATTATATCCAAATTTATGTTT 3329

|||||

Db 568 ATAACAAGTGTGATTGGAAGGCACATGTGAGCATATATGATCCAAATTTATGTTT 627

QY 3330 TTTCCTTTGTTTATATTTTGGGGAAAATTAATAATTTTAAAGGTA 3374  
|||||

Db 628 TTTCCTTTGTTTATATTTTGGGGAAAATTAATAATTTTAAAGGTA 672  
|||||

RESULT 14  
ABL69028/C  
ID ABL69028 standard; DNA: 576 BP.  
XX  
AC ABL69028;  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DE Kidney cancer related gene sequence SEQ ID NO:7365.  
XX  
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200194629-A2.  
XX  
PD 13-DEC-2001.  
XX  
PF 30-MAY-2001; 2001WO-US10838.  
XX  
PR 05-JUN-2000; 2000US-209473P.  
PR 05-JUN-2000; 2000US-209531P.  
PR 18-SEP-2000; 2000US-233133P.  
PR 18-SEP-2000; 2000US-233617P.  
PR 20-SEP-2000; 2000US-234009P.  
PR 20-SEP-2000; 2000US-234034P.  
PR 20-SEP-2000; 2000US-234052P.  
PR 22-SEP-2000; 2000US-234509P.  
PR 22-SEP-2000; 2000US-234567P.  
PR 25-SEP-2000; 2000US-234923P.  
PR 25-SEP-2000; 2000US-234924P.  
PR 25-SEP-2000; 2000US-235077P.  
PR 25-SEP-2000; 2000US-235082P.  
PR 25-SEP-2000; 2000US-235134P.  
PR 25-SEP-2000; 2000US-235280P.  
PR 26-SEP-2000; 2000US-235637P.  
PR 26-SEP-2000; 2000US-235638P.  
PR 27-SEP-2000; 2000US-235711P.  
PR 27-SEP-2000; 2000US-235720P.  
PR 27-SEP-2000; 2000US-235840P.  
PR 27-SEP-2000; 2000US-235863P.  
PR 28-SEP-2000; 2000US-236028P.  
PR 28-SEP-2000; 2000US-236032P.  
PR 28-SEP-2000; 2000US-236033P.  
PR 28-SEP-2000; 2000US-236034P.  
PR 28-SEP-2000; 2000US-236109P.  
PR 28-SEP-2000; 2000US-236111P.  
PR 29-SEP-2000; 2000US-236842P.  
PR 29-SEP-2000; 2000US-236891P.  
PR 02-OCT-2000; 2000US-237172P.  
PR 02-OCT-2000; 2000US-237173P.  
PR 02-OCT-2000; 2000US-237278P.  
PR 02-OCT-2000; 2000US-237294P.  
PR 02-OCT-2000; 2000US-237295P.  
PR 02-OCT-2000; 2000US-237316P.  
PR 03-OCT-2000; 2000US-237425P.  
PR 03-OCT-2000; 2000US-237598P.  
PR 03-OCT-2000; 2000US-237604P.  
PR 03-OCT-2000; 2000US-237606P.  
PR 03-OCT-2000; 2000US-237608P.  
PR 01-NOV-2000; 2000US-244867P.  
PR 01-NOV-2000; 2000US-245084P.  
XX

PA (AVAL-) AVALON PHARM.  
XX  
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX  
XX WPI; 2002-188264/24.  
XX  
XX Screening for anti-neoplastic agent involves exposing cells to a  
XX chemical agent to be tested for anti-neoplastic activity, and  
XX determining a change in expression of a gene of a signature gene set -  
PS Claim 1; SEQ ID 7365; 44pp; English.  
XX  
XX The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a  
CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer.  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilm's tumour.  
XX  
SQ Sequence 576 BP; 171 A; 112 C; 108 G; 182 T; 3 other;  
  
Query Match 14.2%; Score 482.2; DB 24; Length 576;  
Best Local Similarity 97.1%; Pred. No. 2.5e-116;  
Matches 501; Conservative 0; Mismatches 14; Indels 1; Gaps 1;  
  
QY 2860 AAAAGAAGACCATTTTTGAATCTGTT-CAACTGAAACCCCTCAAGATCCCCAAATATATGA 2918  
|||||  
Db 516 AAAGGAGACCCATTTTGAATCTGTCCAACTGAAACCCCTCAAGATCCCCAAATATATGA 457  
  
QY 2919 AGAGACAGTGTCTGTAGCCTTGAGACTAATGAACAAAGAAACCTGCTCTAGTTTACAGGA 2978  
|||||  
Db 456 AGAGACAGTGTGTAGCCCTTGAGACTAATGANCAAGGAAACCTGCTCTAGTTTACAGGA 397  
  
QY 2979 CCATATTTTAGGGTCTGTCCCTCATACCTGTCCATTTGGTGATCTCACAGAGGAGGCCAT 3038  
|||||  
Db 396 CCATATTTTAGGGTCTGTCCCTCATACCTGTCCATTTGGTGATCTCACAGAGGAGGCCAT 337  
  
QY 3039 GCCGCTGAAAAAGGGAAGGAGATTGAAACATTTGATTGCCCTTATCACATGGTCAAGTACCT 3098  
|||||  
Db 336 GCCGCTGAAAAAGGGAAGGAGATTGAAACATTTGATTGCCCTTATCACATGGTCAAGTACCT 277  
  
QY 3099 TGCCCAATAAAGGAAAGCAAAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGA 3158  
|||||  
Db 276 TGCCCAATAAAGGAAAGCAAAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGA 217  
  
QY 3159 AGAGATTTATCTGTATATACACATACTGAAACCAAGTTTAAAGCCCAACTGCACCTGC 3218  
|||||  
Db 216 AGAGATTTATCTGTATATACACATACTGAAACCAAGTTTAAAGCCCAACTGCACCTGC 157  
  
QY 3219 TGATGCATGCCATATAAATTAATGGGTAACCTTTTATCTTTATGATGCTACATAACAAGT 3278  
|||||  
Db 156 TGATGCATGCCATATAAATTAATGGGTAACCTTTTATCTTTATGATGCTACATAACAAGT 97  
  
QY 3279 GTGATTTGGAAGGCACATGTGAGCATATGCATATATGATCCAATTTATGTTTCTTTGT 3338  
|||||  
Db 96 GTGATTTGGAAGGCACATGTGAGCATATGCATATATGATCCAATTTATGTTTCTTTGT 37  
  
QY 3339 TTATATTTTGGGAAAAATTAATAATTTTAAAGGTA 3374  
|||||  
Db 36 TTATATTTTGGGAAAAATTAATAATTTTAAAGGTA 1





Db 1272 TCCAAGCTGGTGA 1284

Search completed: May 12, 2003, 00:39:17  
Job time : 705.486 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.  
  
OM nucleic - nucleic search, using sw model  
  
Run on: May 12, 2003, 00:15:42 ; Search time 181.742 Seconds  
(without alignments)  
6000.498 Million cell updates/sec  
  
Title: US-10-073-333A-3  
Perfect score: 3556  
Sequence: 1 atgctgttcgcgcgcgggg.....ggtaaaaaaaaaaaaaaaa 3556  
  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
  
Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	494	13.9	1129	4	US-09-227-357-40
C 2	47.2	1.3	7218	1	US-08-232-463-14
3	43.2	1.2	1120	3	US-09-030-613-1
4	43.2	1.2	1120	4	US-09-451-905-1
5	42	1.2	1209	4	US-09-105-537-21
6	42	1.2	13613	4	US-09-105-537-3
7	42	1.2	38506	3	US-09-320-878-19
8	41.8	1.2	1743	3	US-08-665-259-20
9	41.8	1.2	1743	3	US-08-762-500-20
10	41.8	1.2	1974	3	US-08-762-500-78
11	41.8	1.2	6803	3	US-08-665-259-19
12	41.8	1.2	6803	3	US-08-762-500-19
13	41.2	1.2	2247	2	US-08-524-828-1
14	41.2	1.2	2247	2	US-08-975-114A-1
C 15	40.8	1.1	16389	4	US-09-741-154-3
16	40.2	1.1	4403765	4	US-09-103-840A-2
17	40.2	1.1	4411529	4	US-09-103-840A-1
18	39.8	1.1	1652	3	US-08-758-662-3
19	39.6	1.1	2834	4	US-09-305-384-6
20	39.6	1.1	6235	4	US-09-305-384-5
21	39.6	1.1	6679	4	US-09-305-384-1
C 22	39.4	1.1	3507	2	US-08-775-009-36
C 23	38.8	1.1	9521	4	US-08-972-218-2
24	38.4	1.1	1283	1	US-07-885-970A-17
25	38.4	1.1	1283	1	US-08-298-687A-17
26	38.4	1.1	1283	1	US-08-530-797-18
27	38.4	1.1	1283	1	US-08-298-829-17

28	38.4	1.1	1283	2	US-08-787-335-18	Sequence 18, Appl
29	38.4	1.1	2371	2	US-08-343-443B-1	Sequence 1, Appli
30	38.2	1.1	319	1	US-07-593-657-14	Sequence 14, Appl
C 31	38.2	1.1	1224	1	US-08-924-847A-1	Sequence 1, Appli
C 32	38.2	1.1	1224	1	US-08-924-847A-3	Sequence 3, Appli
C 33	38.2	1.1	1224	3	US-09-120-052-1	Sequence 1, Appli
C 34	38.2	1.1	1224	3	US-09-120-052-3	Sequence 3, Appli
C 35	38.2	1.1	1415	4	US-09-345-214-11	Sequence 11, Appl
36	38.2	1.1	1534	1	US-08-300-903A-6	Sequence 6, Appli
C 37	38.2	1.1	2008	4	US-09-345-214-12	Sequence 12, Appl
C 38	38.2	1.1	2491	4	US-09-345-214-5	Sequence 5, Appli
39	38.2	1.1	4348	2	US-08-915-868-1	Sequence 1, Appli
C 40	38	1.1	501	2	US-08-997-080-107	Sequence 107, App
C 41	38	1.1	501	2	US-08-997-362-107	Sequence 107, App
C 42	38	1.1	501	4	US-09-095-855-107	Sequence 107, App
C 43	38	1.1	501	4	US-09-324-542-107	Sequence 107, App
C 44	38	1.1	501	4	US-09-205-426-107	Sequence 107, App
C 45	38	1.1	2007	3	US-08-747-221B-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1  
US-09-227-357-40  
; Sequence 40, Application US/09227357  
; Patent No. 6342581  
; GENERAL INFORMATION:  
; APPLICANT: Fischer et al.  
; TITLE OF INVENTION: 123 Human Secreted Proteins  
; FILE REFERENCE: P2010P1  
; CURRENT APPLICATION NUMBER: US/09/227,357  
; CURRENT FILING DATE: 1999-01-08  
; EARLIER APPLICATION NUMBER: PCT/US98/13684  
; EARLIER FILING DATE: 1998-07-07  
; EARLIER APPLICATION NUMBER: 60/051,926  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,793  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,925  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,929  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,803  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,732  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,931  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,932  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,916  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,930  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,918  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,920  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,733  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,795  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,919  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,928  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/055,722  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,723  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,948  
; EARLIER FILING DATE: 1997-08-18





[illegible]

RESULT 3  
US-09-030-613-1  
; Sequence 1, Application US/09030613  
; Patent No. 6083706  
; GENERAL INFORMATION:  
; APPLICANT: Florkiewicz, Robert Z.  
; APPLICANT: Baird, J. Andrew  
; TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:

Query Match 1.2%; Score 43.2; DB 3; Length 1120;  
Best Local Similarity 63.5%; Pred. No. 0.041;  
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

[illegible]

**Qy** 64 CGCGCGGGGCGCTCGCCCGCCTGAGACCCCCGCCTGGATTTCGTG 107  
||| ||||| | - ||| ||| ||| ||| ||| |||  
**Db** 318 CGCGCGGGCTCCAGCGGCTCGGGGATCCGGCCGGGCCCCGCGAG 361

RESULT 4  
US-09-451-905-1  
; Sequence 1, Application US/09451905  
; Patent NO. 6306613  
; GENERAL INFORMATION:

```

; APPLICANT: Robert Z. Florkiewicz
; APPLICANT: Andrew Baird
; APPLICANT: Dale E. Warnock
; TITLE OF INVENTION: MODULATORS OF LEADERLESS PROTEIN EXPORT
; TITLE OF INVENTION: AND METHODS FOR IDENTIFYING AND USING THE SAME
; FILE REFERENCE: 200124.402C4
; CURRENT APPLICATION NUMBER: US/09/451,905
; CURRENT FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1120
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-451-905-1

```

Query Match	1.2%;	Score 43.2;	DB 4;	Length 1120;
Best Local Similarity	63.5%;	Pred. NO. 0.041;		
Matches .66;	Conservative 0;	Mismatches 38;	Indels 0;	Gaps 0;

QY	4	CTGTTCCGGCCCGGGGGCCCGGTACGGGGCAGGGGCTGGGGCGCGCCGGCGAGGCTCCC	63
Db	258	CGGGGCCGTGCCCCGAGCGGGTCGAGGCCGGGGCGGGCGGGGACGGCGGCTCCC	317
QY	64	CGCGCGGGCGCTCGCGCGCCCTGGAGCCCGCGCTGGATTGCTG	107
Db	318	CGCGCGGGCTCAGACGGGCTCGGGATCCGGCGCGGGCCCGCGAG	361

```

RESULT 5
US-09-105-537-21
; Sequence 21, Application US/09105537A
; Patent NO. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-21

```

Query Match	1.2%;	Score 42;	DB 4;	Length 1209;
Best Local Similarity	62.3%;	Pred. No. 0.094;		
Matches 66;	Conservative	0;	Mismatches 40;	Indels 0;
			Gaps	0;

[illegible]

RESULT 6  
US-09-105-537-3  
; Sequence 3, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding me





NUMBER OF SEQUENCES: 83  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENZYME CORPORATION  
STREET: One Mountain Road  
CITY: Framingham  
STATE: Massachusetts  
COUNTRY: United States of America  
ZIP: 01701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/762,500  
FILING DATE: 09-DEC-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/665,259  
FILING DATE: 17-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10469  
FILING DATE: 17-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Dugan, Deborah A.  
REGISTRATION NUMBER: 37,315  
REFERENCE/DOCKET NUMBER: IG5-9.3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (508) 872-8400  
TELEFAX: (508) 872-5415  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1743 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1740  
US-08-762-500-20

Query Match 1.2%; Score 41.8; DB 3; Length 1743;  
Best Local Similarity 59.8%; Pred. NO. 0.14;  
Matches 70; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY	6	GTTCCGCGCCCGGGSCCGGTACGGGGCAGGGGCTGGGGGGCGGCCGGGAGGCCTCCCCG	65
Db	1505	GTAGCGCGCTGTGGTGTCGCCCGCGGGGATGCGGCCCTGCGGCTGCCCGCGCCCTGCTCCCCG	1564
QY	66	CCGCGGGCGCTCGCCCGCCCTGGAGCCCCCGCCCTGGATTGCTGCTGGCGCCTCGCCCG	122
Db	1565	GCCGCGCGCTACTCTCTGCTGGGGGGCGGGCCCTGGAGCCGCGCGCTGGGGGCGCGGGGG	1621

RESULT 10  
US-08-762-500-78  
; Sequence 78, Application US/08762500  
; Patent No. 6030806  
; GENERAL INFORMATION:  
; APPLICANT: Landes, Gregory M.  
; APPLICANT: Burn, Timothy C.  
; APPLICANT: Connors, Timothy D.  
; APPLICANT: Dackowski, William R.  
; APPLICANT: Van Raay, Terence J.  
; APPLICANT: Klinger, Katherine W.  
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,  
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME  
; NUMBER OF SEQUENCES: 83  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENZYME CORPORATION  
; STREET: One Mountain Road  
; CITY: Framingham

STATE: Massachusetts  
COUNTRY: United States of America  
ZIP: 01701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/762,500  
FILING DATE: 09-DEC-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/665,259  
FILING DATE: 17-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10469  
FILING DATE: 17-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Dugan, Deborah A.  
REGISTRATION NUMBER: 37,315  
REFERENCE/DOCKET NUMBER: IG5-9.3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (508) 872-8400  
TELEFAX: (508) 872-5415  
INFORMATION FOR SEQ ID NO: 78:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1974 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-762-500-78

Query Match 1.2%; Score 41.8; DB 3; Length 1974;  
Best Local Similarity 59.8%; Pred. NO. 0.15;  
Matches 70: Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY	6	GTTCGCGCCCGGGSCGGTACGGGCAGGGGCTGGGGGCGGCCGCGGAGGCTCCCCG	65
D <sub>b</sub>	1714	GTAGCGCGCTGTGGTGCCCGCCGGGGATGCGGCCTGCGGCTGCCCGCGCCTGCTCCCCG	1773
QY	66	CCGCGGGCGCTCGCCGCGCCTGGAGCCCCCGCTGGATTGTGCTGGGCGCTCGCCG	122
D <sub>b</sub>	1774	GCCGCGCTACTCTCGTGGGGGGCGGGCCCTGGAGCCGCGGCTGGGGCGCGGGG	1830

RESULT 11  
US-08-665-259-19  
; Sequence 19, Application US/08665259  
; Patent No. 6028173  
; GENERAL INFORMATION:  
; APPLICANT: Landes, Gregory M.  
; APPLICANT: Burn, Timothy C.  
; APPLICANT: Connors, Timothy D.  
; APPLICANT: Dackowski, William R.  
; APPLICANT: Van Raay, Terence J.  
; APPLICANT: Klinger, Katherine W.  
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,  
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME  
; NUMBER OF SEQUENCES: 73  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENZYME CORPORATION  
; STREET: One Mountain Road  
; CITY: Framingham  
; STATE: Massachusetts  
; COUNTRY: United States of America  
; ZIP: 01701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,259
; FILING DATE: 17-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IG5-9.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6803 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-665-259-19

Query Match 1.2%; Score 41.8; DB 3; Length 6803;
Best Local Similarity 59.8%; Pred. No. 0.34;
Matches 70; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 6 GTTCCGCGCCGGGCGGTACGGGGCAGGGGCTGGGGGGCGGGCGGAGGCTCCCGG 65
|| ||||| ||| | ||||| | ||| | ||| ||||| ||||| |||||
Db 5298 GTAGCGCGCTGTGGGTGCCCGCGGGGATGCGGCTGCCCGCGCTGCTCCCGG 5357

QY 66 CCGCGGGCGCTCGCGCCCTGGAGCCCGCTGGATTGCTGCTGGCGCTCGCCGG 122
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5358 GCCGCCGTACCTCCTGCTGGGGGGCGGCGCTGGAGCCGCGGCTGGGGCGCGGGG 5414

RESULT 12
US-08-762-500-19
; Sequence 19, Application US/08762500
; Patent No. 6030806
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/762,500
; FILING DATE: 09-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/665,259
; FILING DATE: 17-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10469
; FILING DATE: 17-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IG5-9.3

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6803 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-762-500-19

Query Match 1.2%; Score 41.8; DB 3; Length 6803;
Best Local Similarity 59.8%; Pred. No. 0.34;
Matches 70; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 6 GTTCCGCGCCGGGCGGTACGGGGCAGGGGCTGGGGGGCGGGCGGAGGCTCCCGG 65
|| ||||| ||| | ||||| | ||| | ||| ||||| ||||| |||||
Db 5298 GTAGCGCGCTGTGGGTGCCCGCGGGGATGCGGCTGCCCGCGCTGCTCCCGG 5357

QY 66 CCGCGGGCGCTCGCGCCCTGGAGCCCGCTGGATTGCTGCTGGCGCTCGCCGG 122
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5358 GCCGCCGTACCTCCTGCTGGGGGGCGGCGCTGGAGCCGCGGCTGGGGCGCGGGG 5414

RESULT 13
US-08-524-828-1
; Sequence 1, Application US/08524828
; Patent No. 5874271
; GENERAL INFORMATION:
; APPLICANT: Atsushi NISHIKAWA et al.
; TITLE OF INVENTION: HUMAN GLYCOSYLTRANSFERASE GENE,
; COMPOUNDS AND METHOD FOR INHIBITING CANCEROUS METASTASIS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/524,828
; FILING DATE: September 7, 1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/107,173
; FILING DATE: August 17, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: James H. Meadows
; REGISTRATION NUMBER: 33,965
; REFERENCE/DOCKET NUMBER: JTF/001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2247 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
US-08-524-828-1

Query Match 1.2%; Score 41.2; DB 2; Length 2247;
Best Local Similarity 62.7%; Pred. No. 0.24;
Matches 64; Conservative 0; Mismatches 38; Indels 0; Gaps 0;







GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.  
  
OM nucleic - nucleic search, using sw model  
Run on: May 11, 2003, 23:19:12 ; Search time 707.514 Seconds  
(without alignments)  
11318.654 Million cell updates/sec

Title: US-10-073-333A-3  
Perfect score: 3556  
Sequence: 1 atgctgttcgcgcgcgggg.....ggtaaaaaaaaaaaaaaaaa 3556

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3556	100.0	3556	22	AAF75055
2	3214	90.4	3390	22	AAF75054
3	1499.6	42.2	1737	22	AAC84891
4	1182.6	33.3	1508	22	AAC84886
5	991.4	27.9	3280	21	AAA95442
6	989.8	27.8	3331	22	AAF89765
7	987.8	27.8	3501	22	AAS21262
8	976.2	27.5	3334	22	AAF28030
9	860.8	24.2	2610	22	AAF89775
					TR16-long receptor
					TR16-short recepto
					Human SEC10 nuclei
					Human SEC5 nucleic
					Human CASB619 prot
					Nucleotide sequenc
					Human cDNA sequenc
					Human TR13 recepto
					Nucleotide sequenc

10	860.8	24.2	2733	22	AAF89774	Nucleotide sequenc
11	761.4	21.4	2554	22	AAF27997	Human TR13 recepto
12	645.4	18.1	672	22	AAS34165	Human cDNA encodin
C 13	504	14.2	576	24	ABL69028	Kidney cancer rela
14	494	13.9	1129	20	AAX27340	Human secreted pro
15	487	13.7	1717	20	AAZ41991	Human endometrium
16	403	11.3	404	22	AAF66437	Novel human polynu
17	381.8	10.7	1149	22	AAF89777	Nucleotide sequenc
18	379.4	10.7	1299	23	ABV23150	Human prostate exp
19	379.4	10.7	1299	23	ABV28991	Human prostate exp
20	285.8	8.0	1119	22	AAH76196	Human seven-transm
21	285.8	8.0	1587	22	AAH76195	Human seven-transm
C 22	244	6.9	244	22	ABA71400	Human foetal liver
C 23	244	6.9	244	22	ABA37625	Probe #16091 for g
C 24	244	6.9	244	22	AAK19720	Human brain expres
C 25	244	6.9	244	22	AAK45740	Human bone marrow
C 26	244	6.9	244	22	AAI51665	Probe #20351 used
C 27	244	6.9	244	24	ABS20017	Human genome-deriv
C 28	234.6	6.6	627	22	AAF89776	Nucleotide sequenc
C 29	218	6.1	371	22	ABA58868	Human foetal liver
C 30	218	6.1	371	22	ABA27759	Probe #6225 for ge
C 31	218	6.1	371	22	AAK07028	Human brain expres
C 32	218	6.1	371	22	AAK32769	Human bone marrow
C 33	218	6.1	371	22	AAI38582	Probe #7268 used t
C 34	218	6.1	371	24	ABS07567	Human genome-deriv
C 35	196.4	5.5	457	22	AAK32324	Human bone marrow
C 36	196.4	5.5	457	24	ABS07104	Human genome-deriv
C 37	196	5.5	196	22	AAK45376	Human bone marrow
C 38	196	5.5	196	24	ABS19637	Human genome-deriv
C 39	182.8	5.1	210	22	ABA70282	Human foetal liver
C 40	182.8	5.1	210	22	ABA36993	Probe #15459 for g
C 41	182.8	5.1	210	22	AAK18511	Human brain expres
C 42	182.8	5.1	210	22	AAK44426	Human bone marrow
C 43	182.8	5.1	210	22	AAI50420	Probe #19106 used
C 44	182.8	5.1	210	24	ABS18661	Human genome-deriv
C 45	177.8	5.0	457	22	ABA57660	Human foetal liver

ALIGNMENTS

RESULT 1  
AAF75055  
ID AAF75055 standard; DNA; 3556 BP.  
XX  
AC AAF75055;  
XX  
DT 10-MAY-2001 (first entry)  
XX  
DE TR16-long receptor DNA.  
XX  
KW TR16 receptor; tumour necrosis factor receptor superfamily;  
KW apoptosis; inflammatory; cancer; immune; neurodegenerative; ds.  
XX  
OS Unidentified.  
XX  
PN WO200112671-A1.  
XX  
PD 22-FEB-2001.  
XX  
PF 10-AUG-2000; 2000WO-US21885.  
XX  
PR 12-AUG-1999; 99US-0148348.  
PR 13-AUG-1999; 99US-0148683.  
PR 13-AUG-1999; 99US-0148870.  
PR 16-AUG-1999; 99US-0148758.  
PR 17-AUG-1999; 99US-0149181.  
PR 18-AUG-1999; 99US-0149453.  
PR 19-AUG-1999; 99US-0149498.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Young PE, Baker KP;

XX	WPI; 2001-138754/14.	
XX		
PT	New nucleic acid molecule encoding a TR16 tumor necrosis factor	
PT	receptor polypeptide, useful for the diagnosis and treatment of cancer,	
PT	autoimmune disorders and cardiovascular diseases -	
XX		
PS	Disclosure; Fig 4; 286pp; English.	
XX		
CC	The present invention relates to a TR16 receptor (tumour necrosis	
CC	factor receptor superfamily). The invention is useful treating	
CC	diseases and disorders associated with the inhibited or increased	
CC	apoptosis. In particular inflammatory diseases, cancers, immune and	
CC	neurodegenerative disorders may be treated.	
XX		
SQ	Sequence 3556 BP; 1043 A; 751 C; 836 G; 926 T; 0 other;	
	Query Match 100.0%; Score 3556; DB 22; Length 3556;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 3556; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 ATGCTGTTCCGCGCCCGGGCGGGTACGGGGCAGGGCTGGGGCGCGCGGAGGCT 60	
DB		
QY	61 CCCCAGCGCGGCGCTCGCGCCCTGGAGCCCCCGCTGGATTGCTGCTGGCGCTCGCC 120	
DB		
QY	121 GGCTGCCAGGCGGCGCTGGGCTGGGACCTGCCCTCCTCCAGCGCGCCGCTTCTCT 180	
DB		
QY	181 TGCCAGGAGAAAGATTATCACATTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGG 240	
DB		
QY	241 AGAGTTGCCATTCCAAATTCGACAGTGGACTGCTCTGGCTGCCCTGACCCAGTGAGAGGC 300	
DB		
QY	301 AAAGATGCACCTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGC 360	
DB		
QY	361 AGTAAGTGTGGTGAAGGCACCTATTCCCTTGGCAGTGGCATCAAATTTGATGAATGSGAT 420	
DB		
QY	421 GAATTCGCGGCAGGATTTCTAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCT 480	
DB		
QY	481 GACAGCAGGCCAGACGGCTGTAAACAACCTCTTCTGGATCCCTCGTGGAAACTACATAGAA 540	
DB		
QY	541 TCTAATCGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGC 600	
DB		
QY	601 TATGTCTTCTTTGAGTACCGGTGTCTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGC 660	
DB		
QY	661 AATGATCAGTGCCAGGAGATGGACACCACTGACAAAGTGGGTAACATTTACAGACAAT 720	
DB		
QY	721 GGAGAATGGGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATCTACTCTGGAGA 780	
DB		
	721 GGAGAATGGGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATCTACTCTGGAGA 780	

QY	781 ACTACAGGCATCCTTATGGGTTCTAAGGCGGTCAAGCCTGTGCTGGTAAAAAATATCACA 840	
DB		
QY	841 ATTGAAGGGTGGCGTACACATCAGAAATGTTTCTTGCAGCAGGCACATTCAGCAAC 900	
DB		
QY	901 AAACCAGGTTCAATCAACTGCCAGTGTGTCCAGAAACACCTATTCTGAGAAAGGAGCC 960	
DB		
QY	961 AAAGAAATGATAAGGTGTAAGACGACTCTCAATTTTCAGGATCCAGTGAGTGTACAGAG 1020	
DB		
QY	1021 CGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATACTCCATGTGATGAAGAAGGA 1080	
DB		
QY	1081 AAGACACAGATAATGTACAAAGTGGATAGGCCCAAAATCTGCCGGGAGGATCTCACAGAT 1140	
DB		
QY	1141 GCTATTAGATTGCCCCCTTCTGGAGAGAAAGGATTGTCCGCCCTTGCNAACCTCGGATT 1200	
DB		
QY	1201 TATAACAATGGATCATCTTCTGCCATCCCTGTCTCCTCCGSAACATTTTCAGATGGAACC 1260	
DB		
QY	1261 AAGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTGAAATATAAATGG 1320	
DB		
QY	1321 TGGAAATGTCTTCCCTGGCAACATGAAAACCTTCCCTGCTCAATGTGGGAAATTCAAAAGTGC 1380	
DB		
QY	1381 GATGGAATGAATGGTTGGGAGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTC 1440	
DB		
QY	1441 GACAATGATTACCTGATCTTAAACTTGATATCCAGGATTAACCCACCAACATCTATG 1500	
DB		
QY	1501 ACTGGAGCCACGGGTTCTGAACTAGGAAGAATAAACAATTTGTCTTTGAGACCCCTCTGT 1560	
DB		
QY	1561 GCTGACTGTGTTTTGTACTTTCATGGTGGATATTAAAGAAAAGTACAAATGTGGTAGAA 1620	
DB		
QY	1621 TCGTGGGGTGGAAACCAAGAAAACAAGCTTACACCCATATCATCTTCAAGAATGCAACT 1680	
DB		
QY	1681 TTTACATTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTCA 1740	
DB		
QY	1741 AATGACATGGTGAAGATTTATTTCTATCACAGCCACTAATGCAGTTGATGGGGTGGCGTCC 1800	
DB		
QY	1801 TCATGCCGTGCCCTGTGCCCTCGGTTCTGAACAGTCCGGTTCATCGTGTGCCCTGCCCT 1860	
DB		
QY	1861 CCAGGCCACTACATTGAGAAAGAAACCAACCAGTGCAGGAATGTCCACCTGACACCTAC 1920	



Db 1861 CCAGGCCACTACATTGAGAAAGAAACCAACCAAGTGCAGGAATGTCACCTTGACACCTAC 1920  
QY 1921 CTGTCCATACATCAGGTCTATGGCAACAGAGCTTGTATTCCATGCGGGCCTGGGAGTAAA 1980  
Db 1921 CTGTCCATACATCAGGTCTATGGCAACAGAGCTTGTATTCCATGCGGGCCTGGGAGTAAA 1980  
QY 1981 AACAAATCAGGACCATTCGGTTTGCTATAGTGACTGCTTTTCTACCATGAAAAGAAAAT 2040  
Db 1981 AACAAATCAGGACCATTCGGTTTGCTATAGTGACTGCTTTTCTACCATGAAAAGAAAAT 2040  
QY 2041 CAGATTTTGCACTATGACTTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGGCCCC 2100  
Db 2041 CAGATTTTGCACTATGACTTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGGCCCC 2100  
QY 2101 AGCTTCACCTCCAAAAGGAACAAAATACTTCCATTTCTTCAATATCAGTTATGTGGCAT 2160  
Db 2101 AGCTTCACCTCCAAAAGGAACAAAATACTTCCATTTCTTCAATATCAGTTATGTGGCAT 2160  
QY 2161 GAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTTTTACAGTAAAAGAAATA 2220  
Db 2161 GAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTTTTACAGTAAAAGAAATA 2220  
QY 2221 GTGGCAGGTCAGATGATTACACAAATTTGGTAGGGGCATTTGTATGCCAGTCAACAATT 2280  
Db 2221 GTGGCAGGTCAGATGATTACACAAATTTGGTAGGGGCATTTGTATGCCAGTCAACAATT 2280  
QY 2281 ATTCTTCTGAAGTAAGGGTTTCCGAGCAGCTTATCATCACAATCCATCATCTGGCA 2340  
Db 2281 ATTCTTCTGAAGTAAGGGTTTCCGAGCAGCTTATCATCACAATCCATCATCTGGCA 2340  
QY 2341 GATACATTCATAGAGTCACAGTTGAAACCAACATGAAATAATTAATATAAAAGAAAGAT 2400  
Db 2341 GATACATTCATAGAGTCACAGTTGAAACCAACATGAAATAATTAATATAAAAGAAAGAT 2400  
QY 2401 ATGTTCCCAGTTCACCAAGCCAAATACCAGATGTGCATTTCTTTTATAAGTCTCTACA 2460  
Db 2401 ATGTTCCCAGTTCACCAAGCCAAATACCAGATGTGCATTTCTTTTATAAGTCTCTACA 2460  
QY 2461 GCAACAACATCTTGTATTAAATGGCCGATCAACTGTGAAAATGAGGTGTAATCCTACT 2520  
Db 2461 GCAACAACATCTTGTATTAAATGGCCGATCAACTGTGAAAATGAGGTGTAATCCTACT 2520  
QY 2521 AAATCTGGAGCAGGAGTGATTTTCAGTCCCCACCAAGTGCCAGCAGGTACCTGTGATGGG 2580  
Db 2521 AAATCTGGAGCAGGAGTGATTTTCAGTCCCCACCAAGTGCCAGCAGGTACCTGTGATGGG 2580  
QY 2581 TGTACGTTCTATTTCCTGTGGGAGAGTGCTGAAGCTTGCCCTCTGTGTACGGAGCATGAC 2640  
Db 2581 TGTACGTTCTATTTCCTGTGGGAGAGTGCTGAAGCTTGCCCTCTGTGTACGGAGCATGAC 2640  
QY 2641 TTCCATGAGATTGAGGAGCCCTGCAAGAGAGAGATTTCAGGAAACCTTGATGTGTGGAAT 2700  
Db 2641 TTCCATGAGATTGAGGAGCCCTGCAAGAGAGAGATTTCAGGAAACCTTGATGTGTGGAAT 2700  
QY 2701 GAACCTAAATGGTGCAATTAAGGAATTTCTTTGCCTGAGAAAAGTTGGCAACCTGTGAA 2760  
Db 2701 GAACCTAAATGGTGCAATTAAGGAATTTCTTTGCCTGAGAAAAGTTGGCAACCTGTGAA 2760  
QY 2761 ACGGTTGACTTTTGGCTGAAGGTGGGAGCCGGTGTTGAGAGCTTTTACTGCCGTTTGTCTG 2820  
Db 2761 ACGGTTGACTTTTGGCTGAAGGTGGGAGCCGGTGTTGAGAGCTTTTACTGCCGTTTGTCTG 2820  
QY 2821 GTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATCAAAAACCTGGAATACAAATATCCAAAG 2880  
Db 2821 GTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATCAAAAACCTGGAATACAAATATCCAAAG 2880  
QY 2881 TTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCCGCTGCAGACAGTTGTGCTATC 2940  
Db 2881 TTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCCGCTGCAGACAGTTGTGCTATC 2940  
QY 2941 ATGGAAGGAGAGATAATGAAGAGGAAGTTGTATATTCCAATAAACAGTCACTACTAGGA 3000

Db 2941 ATGGAAGGAGAGATAATGAAGAGGAAGTTGTATATTCCAATAAAACAGTCACTACTAGGA 3000  
QY 3001 AAACCTCAAAATCTTTGGCAACCAAGGAAAAAGAACACCATTTTGAATCTGTTCAACTGAAA 3060  
Db 3001 AAACCTCAAAATCTTTGGCAACCAAGGAAAAAGAACACCATTTTGAATCTGTTCAACTGAAA 3060  
QY 3061 ACCTCAAGATCCCCAAATATATGAAGAGACAGTGTGTAGCCTTGAGACTAAATGAACAAA 3120  
Db 3061 ACCTCAAGATCCCCAAATATATGAAGAGACAGTGTGTAGCCTTGAGACTAAATGAACAAA 3120  
QY 3121 GAAACCTGCTCTAGTTTACAGGACCATATTTTAGGGTCTGTCTCTCATACCTGTCACATT 3180  
Db 3121 GAAACCTGCTCTAGTTTACAGGACCATATTTTAGGGTCTGTCTCTCATACCTGTCACATT 3180  
QY 3181 GGTGATCTCACAGAGGAGGGCCATGCCGCTGAAAAGGAGGAGATTGAAAACATTTTGATT 3240  
Db 3181 GGTGATCTCACAGAGGAGGGCCATGCCGCTGAAAAGGAGGAGATTGAAAACATTTTGATT 3240  
QY 3241 GCCTTATCACATGGTCAAGTACCTTGCCAAATAAAAGGAAAGCAAAATGATTTGGGTCTCAA 3300  
Db 3241 GCCTTATCACATGGTCAAGTACCTTGCCAAATAAAAGGAAAGCAAAATGATTTGGGTCTCAA 3300  
QY 3301 CTGAAGATGAAGCTCAACTCAGGAAGAGATTATCTGTATATACACATAACTGAAAAACCA 3360  
Db 3301 CTGAAGATGAAGCTCAACTCAGGAAGAGATTATCTGTATATACACATAACTGAAAAACCA 3360  
QY 3361 AGTTTAAAGCCCAACCAATGCACCTGCTGATGCATGCCATATAATTAATGGGTAACCTTTATT 3420  
Db 3361 AGTTTAAAGCCCAACCAATGCACCTGCTGATGCATGCCATATAATTAATGGGTAACCTTTATT 3420  
QY 3421 CTTTATGATGTCTACATAACAAGTGTGATTTGGAAGGCACATGTGAGCATATGCATTATG 3480  
Db 3421 CTTTATGATGTCTACATAACAAGTGTGATTTGGAAGGCACATGTGAGCATATGCATTATG 3480  
QY 3481 ATCCAATTTATGTTTCTTTTGTATATTTTGGGGAAAAATTAATAATTTTAAAGGTA 3540  
Db 3481 ATCCAATTTATGTTTCTTTTGTATATTTTGGGGAAAAATTAATAATTTTAAAGGTA 3540  
QY 3541 AAAAAAIAAAAAAAAA 3556  
Db 3541 AAAAAAIAAAAAAAAA 3556

RESULT 2  
AAF75054

ID AAF75054 standard; DNA; 3390 BP.

XX AAF75054;

XX 10-MAY-2001 (first entry)

XX TR16-short receptor DNA.

XX TR16 recptor; tumour necrosis factor receptor superfamily;  
KW apoptosis; inflammatory; cancer; immune; neurodegenerative; ds.

XX Unidentified.

XX WO200112671-A1.

XX 22-FEB-2001.

XX 10-AUG-2000; 2000WO-US21885.

XX 12-AUG-1999; 99US-0148348.

XX 13-AUG-1999; 99US-0148683.

XX 13-AUG-1999; 99US-0148870.

XX 16-AUG-1999; 99US-0148758.

XX 17-AUG-1999; 99US-0149181.

XX 18-AUG-1999; 99US-0149453.

XX 19-AUG-1999; 99US-0149498.

XX (HUMA-) HUMAN GENOME SCI INC.

```
XX
PI Ruben SM, Young PE, Baker KP;
XX
DR WPI: 2001-138754/14.
XX
PT New nucleic acid molecule encoding a TR16 tumor necrosis factor
PT receptor polypeptide, useful for the diagnosis and treatment of cancer,
PT autoimmune disorders and cardiovascular diseases -
XX
PS Disclosure; Fig 1; 286pp; English.
XX
CC The present invention relates to a TR16 receptor (tumour necrosis
CC factor receptor superfamily). The invention is useful treating
CC diseases and disorders associated with the inhibited or increased
CC apoptosis. In particular inflammatory diseases, cancers, immune and
CC neurodegenerative disorders may be treated.
XX
SQ Sequence 3390 BP; 979 A; 722 C; 801 G; 888 T; 0 other;

Query Match 90.4%; Score 3214; DB 22; Length 3390;
Best Local Similarity 95.3%; Pred. NO. 0;
Matches 3390; Conservative 0; Mismatches 0; Indels 166; Gaps 1;

QY 1 ATGCTGTTCCGCGCCGGGGCCGGTACGGGCGAGGGGCTGGGGCGCGCGGAGGCT 60
DB 1 ATGCTGTTCCGCGCCGGGGCCGGTACGGGCGAGGGGCTGGGGCGCGCGGAGGCT 60

QY 61 CCCC GCCGCGGCGCTCGCCGCCCTGGAGCCCGCCCTGGATTGCTGCTGGCGCTCGCC 120
DB 61 CCCC GCCGCGGCGCTCGCCGCCCTGGAGCCCGCCCTGGATTGCTGCTGGCGCTCGCC 120

QY 121 GGCTGCCAGGCGGCGCTGGCGGCCCTGGAGCCCGCCCTCCCTCCAGCCCGCGCTCCCT 180
DB 121 GGCTGCCAGGCGGCGCTGGCGGCCCTGGAGCCCGCCCTCCCTCCAGCCCGCGCTCCCT 180

QY 181 TGCCAGGAGAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGG 240
DB 181 TGCCAGGAGAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGG 240

QY 241 AGAGTTGCCATTCCAAATCTGCAGTGGAGTGCCTGGCCCTGCCTGACCCAGTGAGAGGC 300
DB 241 AGAGTTGCCATTCCAAATCTGCAGTGGAGTGCCTGGCCCTGCCTGACCCAGTGAGAGGC 300

QY 301 AAAGAATGCATTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCAAGTATGC 360
DB 301 AAAGAATGCATTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCAAGTATGC 360

QY 361 AGTAAGTGGTGAAGGCACCTATTCCCTTGGCAGTGGCATCAAATTTGATGAATGGAT 420
DB 361 AGTAAGTGGTGAAGGCACCTATTCCCTTGGCAGTGGCATCAAATTTGATGAATGGAT 420

QY 421 GAATTGCCGCGCAGGATTTCTAACATCGCAACATTCATGGACACTGTGTGGGCCCTTCT 480
DB 421 GAATTGCCGCGCAGGATTTCTAACATCGCAACATTCATGGACACTGTGTGGGCCCTTCT 480

QY 481 GACAGCAGGCCAGACGGCTGTAAACAACCTTCTTGGATCCCTCGTGGAAACTACATAGAA 540
DB 481 GACAGCAGGCCAGACGGCTGTAAACAACCTTCTTGGATCCCTCGTGGAAACTACATAGAA 540

QY 541 TCTAATCGTGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGC 600
DB 541 TCTAATCGTGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGC 600

QY 601 TATGCTTCTTTGAGTACAGTATGTCGACAACAACATCTTCTTTGAGTCTTTATTCAA 660
DB 601 TATGCTTCTTTGAGTACAGTATGTCGACAACAACATCTTCTTTGAGTCTTTATTCAA 660

QY 661 AATGATCAGTGCCAGGAGATGGACACCACCTGACAAGTGGGTAAACTTACAGACAAT 720
DB 661 AATGATCAGTGCCAGGAGATGGACACCACCTGACAAGTGGGTAAACTTACAGACAAT 720

QY 721 GGAGAATGGGCTCTCATTTCTGTAATGCTGAATCAGGCACAAACATACTACTGGAGA 780
DB 721 GGAGAATGGGCTCTCATTTCTGTAATGCTGAATCAGGCACAAACATACTACTGGAGA 780
```

```
DB 721 GGAGAAATGGGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATACTCTACTGGAGA 780
QY 781 ACTACAGGCATCCTTATGGGTTCTAAGGCGGTCAAGCGTGTGCTGGTAAAAAATATCACA 840
DB 781 ACTACAGGCATCCTTATGGGTTCTAAGGCGGTCAAGCGTGTGCTGGTAAAAAATATCACA 840
QY 841 ATTGAAGGGGTGGGTACACATCAGAATGTTTCTTGCAGGCCAGGCACATTCACCAAC 900
DB 841 ATTGAAGGGGTGGGTACACATCAGAATGTTTCTTGCAGGCCAGGCACATTCACCAAC 900
QY 901 AAACCAGGTTCAATCAACTGCCAGGTGTGTCCCAAGAACACCTATTCTGAGAAAGGAGCC 960
DB 901 AAACCAGGTTCAATCAACTGCCAGGTGTGTCCCAAGAACACCTATTCTGAGAAAGGAGCC 960
QY 961 AAAGAATGTATAAGGTGTAAAGACGACTCTCAATTTTCAGGATCCAGTGAGTACAGAG 1020
DB 961 AAAGAATGTATAAGGTGTAAAGACGACTCTCAATTTTCAGGATCCAGTGAGTACAGAG 1020
QY 1021 CGCCCTCCCTGTACCACAAAAGACTATTTCAGATCCATACCTCCATGTGATGAAGAAGGA 1080
DB 1021 CGCCCTCCCTGTACCACAAAAGACTATTTCAGATCCATACCTCCATGTGATGAAGAAGGA 1080
QY 1081 AAGACACAGATAATGTACAAAGTGGATAGAGCCCAAAAATCTGCCGGAGGATCTCACAGAT 1140
DB 1081 AAGACACAGATAATGTACAAAGTGGATAGAGCCCAAAAATCTGCCGGAGGATCTCACAGAT 1140
QY 1141 GCTATTAGATTGCCCTTCTGGAGAGAGAGGATTGTCCGCCCTTGCAACCCCTGGATTT 1200
DB 1141 GCTATTAGATTGCCCTTCTGGAGAGAGAGGATTGTCCGCCCTTGCAACCCCTGGATTT 1200
QY 1201 TATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 1260
DB 1201 TATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 1260
QY 1261 AAAGAATGTAGACCATGTCCAGCAGGAAACGGAGCCTGCACCTTGGCTTTGAATATAAATGG 1320
DB 1261 AAAGAATGTAGACCATGTCCAGCAGGAAACGGAGCCTGCACCTTGGCTTTGAATATAAATGG 1320
QY 1321 TGGAAATGCTCTTCTGGCAACATGAAACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380
DB 1321 TGGAAATGCTCTTCTGGCAACATGAAACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380
QY 1381 GATGGAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCT 1440
DB 1381 GATGGAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCT 1440
QY 1441 GACAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAACCACCAACATCTATG 1500
DB 1441 GACAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAACCACCAACATCTATG 1500
QY 1501 ACTGGAGCCACGGGTTCTGAACCTAGGAAGATAAATATTTGTCTTTTGAGACCCCTCTGTTCA 1560
DB 1501 ACTGGAGCCACGGGTTCTGAACCTAGGAAGATAAATATTTGTCTTTTGAGACCCCTCTGTTCA 1560
QY 1561 GCTGACTGTGTTTGTACTTTCATGCTGGATATTAAAGAAAAGTACAAAATGTGGTAGAA 1620
DB 1561 GCTGACTGTGTTTGTACTTTCATGCTGGATATTAAAGAAAAGTACAAAATGTGGTAGAA 1620
QY 1621 TCGTGGGTGGAAACCAAGAAAACAAAGCTTACACCCATATCATCTTTCAAGAATGCAACT 1680
DB 1621 TCGTGGGTGGAAACCAAGAAAACAAAGCTTACACCCATATCATCTTTCAAGAATGCAACT 1680
QY 1681 TTTACATTTACATGGGCATTCAGAGAACTAATCAGGGTCAAGGATTAAGACGGTTCATC 1740
DB 1681 TTTACATTTACATGGGCATTCAGAGAACTAATCAGGGTCAAGGATTAAGACGGTTCATC 1740
QY 1741 AATGACATGGTGAAGATTTATCTATCAGACCCACTAATGCAGTTGATGGGTGGCGTCC 1800
DB 1741 AATGACATGGTGAAGATTTATCTATCAGACCCACTAATGCAGTTGATGGGTGGCGTCC 1800
QY 1801 TCATGCCGTGCTGTGCCCTCGGTTCTGACAGTCCGGTTTCATCGTGTGTCGCCCTGCCCT 1860
DB 1801 TCATGCCGTGCTGTGCCCTCGGTTCTGACAGTCCGGTTTCATCGTGTGTCGCCCTGCCCT 1860
```

QY 1861 CCAGGCCACTACATTGAGAAAGAAACCAACCAGTGC AAGGAATGTCCACCTGACACCTAC 1920  
Db 1861 CCAGGCCACTACATTGAGAAAGAAACCAACCAGTGC AAGGAATGTCCACCTGACACCTAC 1920  
QY 1921 CTGTCCATACATCAGGTCTATGGCAAGAGGCTTGT ATTCCATCGGGGCCCTGGGAGTAA 1980  
Db 1921 CTGTCCATACATCAGGTCTATGGCAAGAGGCTTGT ATTCCATCGGGGCCCTGGGAGTAA 1980  
QY 1981 AACAAATCAGGACCATTCGGTTTGTCTATAGTACT GCTTTTTCACCATGAAAAAGAAAAAT 2040  
Db 1981 AACAAATCAGGACCATTCGGTTTGTCTATAGTACT GCTTTTTCACCATGAAAAAGAAAAAT 2040  
QY 2041 CAGATTTTGCACATGACTTTAGCAACCTCAGCAGT GTGGCTCATTAATGAATGGCCCC 2100  
Db 2041 CAGATTTTGCACATGACTTTAGCAACCTCAGCAGT GTGGCTCATTAATGAATGGCCCC 2100  
QY 2101 AGCTTCACCTCCAAAGGAACAAAATACTTCCATT TCTTCAATATCAGTTATGTGGGCAT 2160  
Db 2101 AGCTTCACCTCCAAAGGAACAAAATACTTCCATT TCTTCAATATCAGTTATGTGGGCAT 2160  
QY 2161 GAGGGGAAGAAGATGGCTCTCTGTACCAACAATA TAAACAGACITTTACAGTAAAGAAATA 2220  
Db 2161 GAGGGGAAGAAGATGGCTCTCTGTACCAACAATA TAAACAGACITTTACAGTAAAGAAATA 2220  
QY 2221 GTGGCAGGGTCAGATGATTACACAAATTTGGTAG GGGCATTTGTATGCCAGTCAACAATT 2280  
Db 2221 GTGGCAGGGTCAGATGATTACACAAATTTGGTAG GGGCATTTGTATGCCAGTCAACAATT 2280  
QY 2281 ATTCCCTTCTGAAAGTAAGGGTTTCCGAGCAGCCT TATCATCACAAATCCATCATCTGGCA 2340  
Db 2281 ATTCCCTTCTGAAAGTAAGGGTTTCCGAGCAGCCT TATCATCACAAATCCATCATCTGGCA 2340  
QY 2341 GATACATTTCATAGGAGTCACAGTTGAAACCACAT TTGAAAAATATTAAATATAAAGAGAT 2400  
Db 2341 GATACATTTCATAGGAGTCACAGTTGAAACCACAT TTGAAAAATATTAAATATAAAGAGAT 2400  
QY 2401 ATGTTCCCACTTCCAACAAGCCAAATACCAGAT GTGCATTTCTTTTATAAGCTTCTACA 2460  
Db 2401 ATGTTCCCACTTCCAACAAGCCAAATACCAGAT GTGCATTTCTTTTATAAGCTTCTACA 2460  
QY 2461 GCAACAACATCTTGTATTATTAATGGCCCGATCA AACTGCTGTGAAAAATGAGGTGTAATCCTACT 2520  
Db 2461 GCAACAACATCTTGTATTATTAATGGCCCGATCA AACTGCTGTGAAAAATGAGGTGTAATCCTACT 2520  
QY 2521 AAATCTGGACGAGGAGTGATTTCACTGCCCGATG AAGTGCACCAAGTGCACGAGTACCTGTATGGG 2580  
Db 2521 AAATCTGGACGAGGAGTGATTTCACTGCCCGATG AAGTGCACCAAGTGCACGAGTACCTGTATGGG 2580  
QY 2581 TGTACGTTCTATTTCCTGTGGGAGAGTGCTGAAGC TTGCCCTCTGTGTACGGAGCATGAC 2640  
Db 2581 TGTACGTTCTATTTCCTGTGGGAGAGTGCTGAAGC TTGCCCTCTGTGTACGGAGCATGAC 2640  
QY 2641 TTCCATGAGATTGAGGGAGCCCTGCAAGAGAGGATT TCAGGAAACCTTGATGTGGAAT 2700  
Db 2641 TTCCATGAGATTGAGGGAGCCCTGCAAGAGAGGATT TCAGGAAACCTTGATGTGGAAT 2700  
QY 2701 GAACCTAAATGGTGCAATTAAGGAAATTTCTTTGCC CTGAGAAAAAGTTGGCAACCTGTGAA 2760  
Db 2701 GAACCTAAATGGTGCAATTAAGGAAATTTCTTTGCC CTGAGAAAAAGTTGGCAACCTGTGAA 2760  
QY 2761 ACGGTTGACTTTTGGCTGAAGGTGGGAGCCGGTG TGCGGAGCTTTTACTGCCGTTTGTCTG 2820  
Db 2761 ACGGTTGACTTTTGGCTGAAGGTGGGAGCCGGTG TGCGGAGCTTTTACTGCCGTTTGTCTG 2820  
QY 2821 GTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATC AAAAACTGGAATACAAATATTCCAAG 2880  
Db 2821 GTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATC AAAAACTGGAATACAAATATTCCAAG 2880  
QY 2881 TTAGTAATGACGACTAACTCAAAAGAGTGTGAACCT CCCGGCTGCAGACAGTTGTGCTATC 2940  
Db 2859 ----- 2858

QY 2941 ATGGAAGGAGAAAGATAATGAAGAGGAAGTTGTAT ATTCCAATAAAACAGTCACTACTAGGA 3000  
Db 2859 ----- 2858  
QY 3001 AAACCTCAAAATCTTTGGCAACCAAGGAAAAAG AAGAACCATTTTGAATCTGTTCAACTGAAA 3060  
Db 2859 -----GAAAAAGAAAGAACCATTTTGAATCT GTTCAACTGAAA 2894  
QY 3061 ACCTCAAGATCCCCAAATATATGAAGAGACAGTG CTGTAGCCCTTGAGACTAATGAACAAA 3120  
Db 2895 ACCTCAAGATCCCCAAATATATGAAGAGACAGTG CTGTAGCCCTTGAGACTAATGAACAAA 2954  
QY 3121 GAAACCTGCTCTAGTTTACAGGACCATATTTTAG GGTCTGTCTCATACCTGTACACATT 3180  
Db 2955 GAAACCTGCTCTAGTTTACAGGACCATATTTTAG GGTCTGTCTCATACCTGTACACATT 3014  
QY 3181 GGTGATCTCACAGAGGAGGGCCATGCCGCTGAAA AAGGAGAGATTGAAACATTTTGATT 3240  
Db 3015 GGTGATCTCACAGAGGAGGGCCATGCCGCTGAAA AAGGAGAGATTGAAACATTTTGATT 3074  
QY 3241 GCCTTATCACATGGTCAAGTACCTTGCCAAATAAA AGGAAAGCAAAATGATTTGGGTCTCAA 3300  
Db 3075 GCCTTATCACATGGTCAAGTACCTTGCCAAATAAA AGGAAAGCAAAATGATTTGGGTCTCAA 3134  
QY 3301 CTGAAGATGAAGCTCAACTCAGGAAGAGATTATCT GTATATACACATAACTGAAAACCA 3360  
Db 3135 CTGAAGATGAAGCTCAACTCAGGAAGAGATTATCT GTATATACACATAACTGAAAACCA 3194  
QY 3361 AGTTTAAGCCCAACCAATGCACCTGCTGATGCC ATGCCATATAATTAATGGGTAACTTTTATT 3420  
Db 3195 AGTTTAAGCCCAACCAATGCACCTGCTGATGCC ATGCCATATAATTAATGGGTAACTTTTATT 3254  
QY 3421 CTTTATGATGTCTACATAACAAGTGTGATTTGGA AGGCACACATGTGAGCATATGCATTATG 3480  
Db 3255 CTTTATGATGTCTACATAACAAGTGTGATTTGGA AGGCACACATGTGAGCATATGCATTATG 3314  
QY 3481 ATCCAAATTTATGTTTTTCTTTGTTTATATTTTG GGGGAAAAATTAATAATTTTTTAAGGTA 3540  
Db 3315 ATCCAAATTTATGTTTTTCTTTGTTTATATTTTG GGGGAAAAATTAATAATTTTTTAAGGTA 3374  
QY 3541 AAAAAAATAAAAAAAAAA 3556  
Db 3375 AAAAAAATAAAAAAAAAA 3390

RESULT 3  
AAC84891  
ID AAC84891 standard; cDNA; 1737 BP.  
XX AAC84891;  
XX  
DT 20-APR-2001 (first entry)  
XX Human SEC10 nucleic acid sequence (clone ID 1795045.0.77).  
DE  
XX  
KW SECX; cytostatic; gynecological; gene therapy; screening assay; human;  
KW chromosomal mapping; forensic biology; cell proliferation; cancer;  
KW cell differentiation; immune associated disorder; gestational disease;  
KW SEC10; ss.  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 296..1690  
FT /\*tag= a  
FT /product= "SEC10"  
XX  
PN WO200078802-A2.  
XX  
PD 28-DEC-2000.  
XX  
PF 23-JUN-2000; 2000WO-US17328.  
XX







Db 1430 TTCCCAGAGAACTAATTCAGGGTCCCAAGATAATAGACGGTTCNCNCAT 1477

RESULT 5  
ID AAA95442  
XX AAA95442 standard; cDNA; 3280 BP.  
XX  
DT 12-FEB-2001 (first entry)  
XX  
DE Human CASB619 protein coding sequence #1.  
XX  
KW Human; CASB619; cancer; autoimmune disease; immunogen; vaccine;  
KW epitope; ss.  
XX Homo sapiens.  
OS  
XX  
FH Key Location/Qualifiers  
XX CDS 1..3342  
FT /\*tag= a  
FT /product= "CASB619"  
XX

PN WO200058460-A2.

XX 05-OCT-2000.

PF 20-MAR-2000; 2000WO-EP02478.

XX 26-MAR-1999; 99GB-0007113.

PR 25-SEP-1999; 99GB-0022858.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX

PI Bruck CEM, Cassart J, Coche T, Vinals De Bassols YC;

XX WPI; 2000-664923/64.

DR P-PSDB; AAB26179.

PT Novel CASB619 polypeptides useful for diagnosing, and as vaccines for  
PT prophylactic and therapeutic treatment of, cancers, particularly  
PT ovarian and colon carcinoma, and autoimmune diseases -

XX Claim 13; Page 53-54; 68pp; English.

CC The present sequence comprises the human CASB619 coding sequence. This  
CC protein is thought to be specifically or over-expressed in tumour cells,  
CC and so can be used as a target for antigen-specific immune responses  
CC which can cause destruction of the tumour cell. In addition, the protein  
CC and gene can be used in cancer diagnosis, in the treatment of autoimmune  
CC diseases and in vaccines against cancer and autoimmune disease. The  
CC invention provides a number of epitopes derived from the protein which  
CC can be used as immunogens.

XX SQ Sequence 3280 BP; 810 A; 930 C; 797 G; 743 T; 0 other;

Query Match 27.9%; Score 991.4; DB 21; Length 3280;  
Best Local Similarity 60.6%; Pred. No. 6.1e-250;  
Matches 1758; Conservative 0; Mismatches 1116; Indels 27; Gaps 7;

QY 171 GCTTCCTCCTGCCAGGAGAAAGATTATCAGTTTGAATATACGGAAATGTGATAGCAGTGG 230

Db 135 GCTTCATGCCGCAAGAGAGTCTGAGTACCAGTATGAGTACACGGCGTGTGACAGCAGGG 194

QY 231 CTCCAGGTGGAGAGTTGCCATTCCAAATCTGCAGTGGAGTGGCTGGCCCTGCCTGACCC 290

Db 195 TTCCAGGTGGAGGTGCGCGTGGCCGATACCCCGGCCCTGTGCACACAGCCTGCCTGACCC 254

QY 291 AGTGAGAGGCAAGAAGATGCACATTCTCCTGTGCTTCTGGAGAGATFCTAGAAATGAAGAA 350

Db 255 CGTCAAGGSCACCGAGTGTCTCCTCTCTGCAACGCCGGGAGTTTCTGGATATGAAGGA 314

QY 351 CCAGGTATGCAGTAAGTGTGGTGAAGGGCAGCCTATTTCCTTGGGCGAGTGGCATCAAAATTTGA 410  
Db 315 CCAGTCATGTAAGCCCATGCGCTGAGGGCGCGCTACTCCCTCGGCACAGGCATTCGGTTTGA 374  
QY 411 TGAATGGGATGAATTCCGGCAGGATTTTCTTAACATCGCAACATTCATGGACACATGTGGT 470  
Db 375 TGAGTGGGATGAGTGGCCCATGGCTTTTGGCAGCCTCTCAGCCCAACATGGAGCTGGATGA 434  
QY 471 GGGCCCTTCGACAGCAGGCCGAGACGGCTGTAAACAACTCTTCTTGGATCCCTCGTGGAAA 530  
Db 435 CAGTGTCTGCTGAGTCCA--CCGGGAAGTGTACTTCTGTCCTCAAGTGGGTCCCGGGGCGA 491  
QY 531 CTACATAGAAATCTAATCGTGTGATGACTGCACGGTGTCTTTTGATCTATGCTGTCCACCTTAA 590  
Db 492 CTACATCGCCTCCCAACACAGGACGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 551  
QY 591 GAAGTCAGGCTATGTCTTTTGTAGTACCAGTATGTGACAAACAACTCTTCTTTGAGTTT 650  
Db 552 GCAATCTGGCAGCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTTT 611  
QY 651 CTTTATTCAAAATGATCAGTCCAGGAGATGGACACCCACTGACAACTGGGTAAAACT 710  
Db 612 TTTTCGTTCAGAAATGACCAAGTCCAGGCCCAATGCAGATGACTC--CAGGTGGATGAAGAC 668  
QY 711 TACAGACAAATGGAGAATGGGGCTCTCATTTCTGTAAATGCTGAAATCAGGCACAAACATACT 770  
Db 669 CACAGAGAA--AGGATGGAAATTCACACAGTGTGGAGCTAAATCGAGGCAATTAATGTCTT 725  
QY 771 CTACTGGAGAACTACAGGCTATCCTTTATGGGTTCTAAGGGCGGTCAAGCCTGTGCTGTTAA 830  
Db 726 CTATTGGAGAACCAACAGCCTTCTCAGTATGGACCAAAAGTACCCCAAGCCTGTGCTGGTGA 785  
QY 831 AATATACAAATGAAGGGGTGGCGTACACATCAGAAATGTTTTCTTGCAGGCCAGGCAC 890  
Db 786 AAACATTCGCAATAACAGGGGTGGCGTACACTTCAGAAATGCTTCCCTGCAAACTCTTATTTCAA 845  
QY 891 ATTACAGCAACAAACAGGTTTCATTCAACTGCCAGGTGTGTCCCAGAAACACCTATTTCTGA 950  
Db 846 GTATGCAGACAAAGCAGGGCTCCTCTTTCTGCAAACTTTGCCCAAGCAACTCTTATTTCAA 905  
QY 951 GAAAGGAGCCAAAGAAATGTATAAGGTGTAA--AGACGACTCTCAATTTTCAGGATCCAG 1007  
Db 906 TAAAGGAGAAACTTCTTGGCCACAGTGTGACCCCTGACAAATACTCAGAGAAAGGATCTTTC 965  
QY 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATCTCCATG 1067  
Db 966 TTCTCTGTAACTGCGCCCGAGCTTGCACAGACAAAGATTATTTCTACACACACACGGCCTG 1025  
QY 1068 TGATGAAGAGGAAAGACACAGATTAATGTACAAAGTGGATAGAGCCCAAAATCTGCCGGGA 1127  
Db 1026 CGATGCCACGGAGAGACACAACTCATGTACAAATGGGCCAAGCCGAAATCTGTAGCGA 1085  
QY 1128 GGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAGAAGGATGTCCGCTTG 1187  
Db 1086 GGACCTTGAGGGGGCAGTGAAGCTGCCCTGCTGCTGTGGTGTGAAGACCCACTGCCACCTG 1145  
QY 1188 CAACCCCTGGATTTTATAACAATGGATCATCTTTTGGCATCCCTGCTCTCTGGAACATTT 1247  
Db 1146 CAACCCAGGCTTCTTCAAAACCAACACAGCAGCCTGCCAGCCCTGCCCATATGTTTCTTA 1205  
QY 1248 TTTCAGATGGAACCAAAAGAAATGTAGACCAATGTCCAGCAGGAACGGAGCTGCACTTGGCTT 1307  
Db 1206 CTCCAATGG--CTCAGACTGTACCCCGTGGCTGCGAGGACTGAACCTGCTGTGGGATTT 1262  
QY 1308 TGAATATAAATGGTGAATGTCTCTTCTGGCAACATGAAAACTTCTCTCTCAATGTTGG 1367  
Db 1263 TGAATACAAATGGTGAACACAGCTGCCCAACAAACATGGAACAGCCGTTCTCAGTGGGAT 1322  
QY 1368 GAATTCAGAGTGGGATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427  
Db 1323 CAACTTCGAGTACAAAGGGCATGACAGGGTGGGAGGTGGCTGGTGTATCATATTTACACAGC 1382  
QY 1428 GGCTGGAGGTCTGTGACAAATGATTAACCTGATCTTAAACTTGCATATCCAGGATTTAAACC 1487



Db 1383 TGCTGGAGCCTCAGACAATGACTTCATGATTTCTCACTCTGGTTGTGCCAGGATTTAGACC 1442  
QY 1488 ACCAACATCTATGACTGGAGCCAC---GGGTTCTGAACTAGGAGAATAACATTTGTCTT 1544  
Db 1443 TCCGCAGTCGGTGATGGCAGACACAGAGAATAAAGAGGTGGCAGAAATCACATTTGTCTT 1502  
QY 1545 TGAGACCCCTCTGTTCAGCTGACTGTGTTTGTACTTTCATGGTGGATATTAATAGAAAAAG 1604  
Db 1503 TGAGACCCCTCTGTTCTGTGAACCTGTGAGCTCTACTTCATGGTGGGTGTAATTTCTAGGAC 1562  
QY 1605 TACAAATGTGGTAGAATCGTGGGTGGAAACCAAGAAAAAACAAGCTTACACCCATATCAT 1664  
Db 1563 CAACACTCCTGTGGAGACGTGGAAGGTTCCAAAGGCCAAACAGTCTATACCTACATCAT 1622  
QY 1665 CTTCAAGAATGCAACTTTTACATTTACATGGGCGCATTTCCAGAGAACTAATCAGGGTCAAGA 1724  
Db 1623 TGAGGAGAACACTACCACGAGCTTCACCTGGGCGCTTCCAGAGGACCACTTTTCATGAGGC 1682  
QY 1725 TAATAGACGGTTTCATCAATGACATGGTGAAGATTTATTTCTATCACAGCCCACTAATGCAGT 1784  
Db 1683 AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCAACCAATGTTAT 1742  
QY 1785 TGATGGGGTGGCGTCCCTCATGCCGTGCTGTGCCCTCGGTTCTGACAGTCGGGTTTCATC 1844  
Db 1743 GAATGGCGTGGCCTCCTACTGCCGTCCCTGTGCCCTAGAAAGCCTCTGATGTGGGCTCCTC 1802  
QY 1845 GTGTGTCCCTCGCCCTCCAGGCCCACTACATTTGAGAAAGAAACCAACCAGTGCAAGGAATG 1904  
Db 1803 CTGCACCTCTGTCTGCTGCTGTTACTATATTGACCGAGATTCAGGAACCTGCCACTCCTG 1862  
QY 1905 TCCACCTGACACCTACCTGTCCATACATCAGGTCTATGGCAAAGAGCCTGTATTCATG 1964  
Db 1863 CCCCCCTAACACAATTCTGAAAGCCCAACAGCCTTATGGTGTCCAGGCGTGTGCCCTG 1922  
QY 1965 CGGGCCTGGGAGTAAAAACAATCAGGACCATTCGGTTTGCTATAGTACTGCTTTTCTA 2024  
Db 1923 TGGTCCAGGGACCAAGAACCAAGATCCACTCTCTGTGCTACAAATGATGACCTTTCTC 1982  
QY 2025 CCATGAAAAAGAAATCAGATTTTGCACATGACTTTTAGCAACCTCAGCAGTGTGGGCTC 2084  
Db 1983 ACGCAACACTCCAACACGAGACTTTCAACTACAACTTCTCCGCTTTGGCAAAACACCGTCAC 2042  
QY 2085 ATTAATGAATGSCCCAGCTTCACCTCCAAGGAACAAAATACITTCACATTTCTTCAATAT 2144  
Db 2043 TCTTGCTGGAGGCCAAGCTTCACTTCCAAGGGTTGAAATACTTCCATCACTTTACCCCT 2102  
QY 2145 CAGTTTATGTGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTT 2204  
Db 2103 CAGTCTCTGTGGAACACGAGGGTAGGAAATGTCTGTGTGCACCGACAAATGTCACTGACCT 2162  
QY 2205 TACAGTAAAGAAATAGTGGCAGGGTCAGATGATTACACAAAATTTGGTAGGGGCATTTGT 2264  
Db 2163 CCGGATTCCTGAGGGTGAGTCAGGG-----TTCTCCAAATCTATCACAGCCTACGT 2213  
QY 2265 ATGCCAGTCAACAATATTCCTTCTGAAAAGTAAAGGGTTTCCGAGCAGCCTTATCATCACA 2324  
Db 2214 CTGCCAGGCAGTCATCATCCCCCAGAGGTGACAGGCTACAAGGCCGGGTTTCCCTCACA 2273  
QY 2325 ATCCATCATTTCTGGCAGATACATTTCATAGGAGTCAAGTTGAAACCAACATTTGAAAAATAT 2384  
Db 2274 GCCTGTCAGCCTTGCTGATCGACTTATTGGGGTGACAACAGATATGACTCTGGATGGAAT 2333  
QY 2385 TAATATAAAGAGAAGATATGTTCCAGTTTCCAAAGCCAAATACCAGATGTGCATTTCTT 2444  
Db 2334 CACCTCCCCAGCTGAACCTTTTCCACCTGGAGTCCTTGGGAATACCGGACGTGATCTTCTT 2393  
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAATGGCCGATCAACTGCTGTGAAAAAT 2504  
Db 2394 TTATAGGTCCAATGATGTGACCCAGTCTGCACTTCTGGGAGATCAACCACCATCCGCGT 2453  
QY 2505 GAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCCCAGCAAGTCCCCAGC 2564

Db 2454 CAGGTGCAGTCCACAGAAAACTGTCCCTGGAAGTTTGTGCTGCCAGGAACGTGCTCAGA 2513  
QY 2565 AGGTACCTGTGATGGGTGTACGTTTCTATTTCTGTGGGAGAGTGTGTAAGCTTGCCTCT 2624  
Db 2514 TGGGACCTGTGATGGTGTCAACTTCCACTTCTGTGGGAGAGCGCGGCTGCTTGCCCGCT 2573  
QY 2625 GTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCGCTGCAAGAGAGGATTTTCAGGAAAC 2684  
Db 2574 CTGCTCAGTGGCTGACTACCATGCTATCGTCAGCAGCTGTGTGGCTGGGATCCAGAAGAC 2633  
QY 2685 CTTGTATGTGTGAATGAACCTAAATGCTGCATTAAAGGAATTTCTTTCCTTGAGAAAAA 2744  
Db 2634 TACTTACGTGTGGCAGAGAACCCCAAGCTATGCTCTGGTGCAATTTCTGCTGAGCAGAG 2693  
QY 2745 GTTGGCAACCTGTGAAACGGTTGACITTTTGGCTGAAGGTGGGAGCGGCTGTGGAGCTTT 2804  
Db 2694 AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG 2753  
QY 2805 TACTGCCGTTTGTGCTGGTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATCAAAAACTGGA 2864  
Db 2754 TACTGCCATCTGCTCACCGTCTTGACCTGCTACTTTTGGAAAAAGAAATCAAAAACTAGA 2813  
QY 2865 ATACAAATATTCCAAAGTTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCCGGCTGC 2924  
Db 2814 GTACAAGTACTCCAAGCTGGTGATGAATGCTACTCTCAAGGACTGTGACCTGCCAGCAGC 2873  
QY 2925 AGACAGTTGTGCTATCATGGAAGGAGAGATAAATGAAGAGAGTTGTATATTTCCAATAA 2984  
Db 2874 TGACAGCTGCCCATCATGGAAGCGGAGGATGTAGAGGACGACCTCATCTTTACCAGCAA 2933  
QY 2985 ACAGTCACCTACTAGGAAAACTCAAATCTTTGGCAACCAAGGAAAAAGAGACCATTTTGA 3044  
Db 2934 GAAGTCACCTCTTTGGGAAGATCAAATCATTTACCTCCAAGAGGACTCCTGTGATGGATTGA 2993  
QY 3045 ATCTGTTCAACTGAAAAACCTC 3065  
Db 2994 CTCAGTGCCCGCTGAAGACATC 3014

RESULT 6  
AAF89765  
ID AAF89765 standard; cdna; 3331 BP.  
XX  
AC AAF89765;  
XX  
DT 23-JUL-2001 (first entry)  
XX  
DE Nucleotide sequence of a human protein expressed in tumour cells.  
XX  
KW Tumour cell; immunological disease; autoimmune disease; cancer;  
KW infection; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 45..3086  
FT /\*tag= a  
FT /\*product= "tumour expressed protein"  
FT sig\_peptide 45..167  
FT /\*tag= b  
XX  
PN WO200131003-A1.  
XX  
PD 03-MAY-2001.  
XX  
PF 30-OCT-2000; 2000WO-FR03032.  
XX  
PR 29-OCT-1999; 99FR-0013629.  
XX  
PA (FABR ) FABRE MEDICAMENT SA PIERRE.  
XX  
PI Delneste Y, Magistrelli.G, Jeannin P, Bonnefoy J;  
XX

DR WPI; 2001-328651/34.  
DR P-PSDB; AAB83845.  
XX  
PT New nucleic acid, expressed in tumours and lymphoid tissue is useful for  
PT identifying agents for treating tumours and autoimmune disease -  
XX  
PS Claim 1; Page 43-48; 85pp; French.  
XX  
CC The present sequence encodes a human protein expressed in tumour cells.  
CC The polynucleotide is useful for screening cDNA/genomic DNA banks and  
CC for cloning isolated DNA; identifying mutant forms of the gene that  
CC encodes a human protein, where the mutations are associated with  
CC abnormal gene expression, or promoters and regulators of the gene,  
CC particularly for diagnosis; for recombinant expression of the derived  
CC protein; as probes and primers for detection and amplification; and  
CC as antisense therapeutics. The tumour expressed protein is useful for  
CC raising specific antibodies and to screen agents that modulate its  
CC activity, bind to it or interact with it. These agents are potentially  
CC useful for treatment or prevention of diseases associated with abnormal  
CC expression/activity of the protein, particularly immunological diseases  
CC (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic  
CC infections.  
XX  
SQ Sequence 3331 BP; 822 A; 945 C; 813 G; 751 T; 0 other;  
  
Query Match 27.8%; Score 989.8; DB 22; Length 3331;  
Best Local Similarity 60.6%; Pred. No. 1.6e-249;  
Matches 1757; Conservative 0; Mismatches 1117; Indels 27; Gaps 7;  
  
QY 171 GCTTCCTCCTGCCAGGAGAAAGATATATACACTTTGAATATACGGAATGTGATAGCAGTGG 230  
DB 179 GCTTCATGCTGCAAGAGTCTGAGTACCACATATGAGTACACGGCGTGTGACAGCACGGG-238  
  
QY 231 CTCACGGTGGAGAGTTGCCATTCCAAATTCTGCAGTGGACTGCTCTGGCCCTGCCTGACCC 290  
DB 239 TTCCAGGTGGAGGGTGCCTGTCGCCATACCCCGGCGCTGTGCACACGCTGCTGACCC 298  
  
QY 291 AGTGAGAGGCAAGAATGCACCTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAA 350  
DB 299 CGTCAAGGCGACCGAGTGTCTCTCTGCAACGCCGGGAGTTTCTGGATATGAAGGA 358  
  
QY 351 CCAGGTATGCAGTAAGTGTGGTGAAGGCACCTATTTCTTGGGAGTGGCATCAAAATTTGA 410  
DB 359 CCAGTCAATGAAGCCATGCGCTGAGGGCGCTACTCCCTCGGCACAGGCATTCGGTTTGA 418  
  
QY 411 TGAATGGGATGAATTCCCGGCAGGATTTCTTAACATCGCAACATTCATGGACACTGTGGT 470  
DB 419 TGAGTGGGATGAGCTGCCCATGGCTTTGCCAGCCTCTCAGCCAACATGGAGCTGGATGA 478  
  
QY 471 GGGCCCTCTGCACAGCAGGCCGCTGTAAACAACCTCTTCTTGGATCCCTCGTGGAAA 530  
DB 479 CAGTGTCTGAGTCCA--CCGGGAAGTGTACTTGGTCCAAAGTGGGTTCCTCCGGGGCGA 535  
  
QY 531 CTACATAGAAATCTAATCGTGATGACTGCACGGGTGCTTTTGATCTATGCTGTCACCTTAA 590  
DB 536 CTACATCGCTCCAAACACGGACGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 595  
  
QY 591 GAAGTCAGGCTATGCTCTTTTGTAGTACCAGTATGTCGACACAAACATCTTCTTTGAGTT 650  
DB 596 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCTTTTGAGTT 655  
  
QY 651 CTTTATTCAAATGATCAGTGCCAGGAGATGGACACCACCTATGACAAGTGGGTAAACT 710  
DB 656 TTTCGTTTCAGAAATGACCAAGTGCAGGCCCAATGCAGATGACTC--CAGGTGGATGAAGAC 712  
  
QY 711 TACAGACAATGGAGAATGGGGCTCTCATTTCTGTAATGCTGAATTCAGGCACAAACATACT 770  
DB 713 CACAGAGAA---AGGATGGGAATTCACAGTGTGGAGCTAAATCAGGGCAATAATGTCT 769  
  
QY 771 CTACTGGAGAACTACAGGCATCCTTATGGGTTCTAAGGGGCTCAAGCCTGTGCTGGTAA 830  
DB 770 CTATTGGAGAACACAGCCTTCTCAGTATGGACCAAGTACCCCAAGCCTGTGCTGGTGA 829

QY 831 AAATATCACAATTTGAAGGGTGGCGGTACACATCAGAAATGTTTCTCTTGAAGCCAGGCAC 890  
DB 830 AAACATTTGCCATAACAGGGGTGGCCTACACTTCAGAAATGCTTCCCCTGCAAAACCTGGCAC 889  
  
QY 891 ATTTCAGCAACAACAAACCAGGTTTCATTTCAACTGCCAGGTGTGTCCCGAGAAACACCTATTCTGA 950  
DB 890 GTATGCAGACAAGCAGGGCTCTCTTTCTGCAAAACTTTGCCAGGCCAACTCTTATTCAAA 949  
  
QY 951 GAAAGGAGCCCAAGAAATGTATAAGGTGTAA--AGACGACTCTCAATTTTCAGGATCCAG 1007  
DB 950 TAAAGGAGAAACTTCTTGCACACAGTGTGACCCCTGACAAATACTCAGAGAAAGGATCTTC 1009  
  
QY 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCCATCTCCATG 1067  
DB 1010 TTCTCTTAACGTGCGCCAGCTTGCACAGACAAAGATTTATTCTACACACACACGCGCTG 1069  
  
QY 1068 TGATGAAGAAAGGAAAGACACAGATAATGTACAAGTGGATAGAGCCCCAAAAATCTGCCGGA 1127  
DB 1070 CGATGCCAACGGAGAGACACAACTCATGTACAAATGGGCCAACCGGAAATCTGTAGCGA 1129  
  
QY 1128 GGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAAGAGATTGTCCCGCTTG 1187  
DB 1130 GGACCTTGAGGGGGCAGTGAAGCTGCCTGCCTCTGGTGTGAAGACCCCACTGCCCACTTG 1189  
  
QY 1188 CAACCCCTGGATTTTTATAACAATGGATCATCTTCTTGGCATCCCTGTCTCTCTCTGGAACATTT 1247  
DB 1190 CAACCCAGGCTTCTTCAAAACCAACAACACAGCACCTGCCAGCCCTGCCCATATGGTCCCTA 1249  
  
QY 1248 TTTCAGATGGAACCAAGAATGTAGACCATGTCCAGCAGGAACGGAGCCCTGCACCTTGGCTTT 1307  
DB 1250 CTCCAATGG--CTCAGACTGTACCCCGCTGCCCTGCAGGGACTGAAACCTGCTGTGGGATTT 1306  
  
QY 1308 TGAATATAAATGGTGGAAATGTCTTCTTGGCAACATGAAAACTTCTGCTTCAATGTTGG 1367  
DB 1307 TGAATACAAATGGTGGAAACACGCTGCCCAACAACATGGAACGACGACCGTCTCAGTGGGAT 1366  
  
QY 1368 GAATTCAAAAGTCCGATGGAATGAATGGTGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427  
DB 1367 CAACCTCGAGTACAAGGGCATGACAGGCTGGGAGGTGGCTGGTGTATCACATTTACACAGC 1426  
  
QY 1428 GGCTGGAGGTTCTGACAAATGATTTACCTGATCTTAAACTTTCGATATCCAGGAATTAACCC 1487  
DB 1427 TGCTGGAGCCTCAGACAATGACTTTCATGATTTCTCACTCTGCTGTGTCAGAGATTTAGACC 1486  
  
QY 1488 ACCAACATCTATGACTGGAGCCAC--GGGTCTTGAACCTAGGAAGAAATAACATTTGCTTT 1544  
DB 1487 TCCGCGAGTGGTGTATGGCAGACACAGAGAATAAAGAGGTGGCCAGAAATCACATTTGTTCTT 1546  
  
QY 1545 TGAGACCCCTCTGTTTCAGCTGACTGTGTTTTGTACTTTCATGTTGGATATTAATAGAAAAG 1604  
DB 1547 TGAGACCCCTCTGTTCTGTGAACCTGTGAGCTCTACTTTCATGGTGGTGTGAATTTCTAGGAC 1606  
  
QY 1605 TACAAATGTGGTAGAATCGTGGGTGGAAACCAAGAAAGAAAGCTTTACACCCCATATTCAT 1664  
DB 1607 CAACACTCCTGTGGAGACGTGGAAGGTTCCAAAGGCCAAACAGTCTCTATACCTACATTCAT 1666  
  
QY 1665 CTTTCAAGAATGCAACTTTTACATTTACATGGGCATTTCCAGAGAACTAAATCAGGGTCAAGA 1724  
DB 1667 TGAGGAGAAACACTACCACGAGCTTTCACCTGGCCCTTCCAGAGGACCACTTTTCATGAGGC 1726  
  
QY 1725 TAATAGACGGTTTCATCAATGACATGGTGAAGATTTATTTCTATCAGACGCCACTTAATGCAGT 1784  
DB 1727 AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTACCAATGTTAT 1786  
  
QY 1785 TGATGGGGTGGCGTCTCTCATGCCGTGCCCTGTGCCCCCTCGGTTCTGAAACAGTCCGGTTCAATC 1844  
DB 1787 GAATGGCGTGGCCCTCCTACTGCGCGTCCCTGTGCCCCCTAGAAAGCTCTGATGTGGGCTCCTC 1846  
  
QY 1845 GTGTGTCCCTTCCCTCCAGGCCACTTACATTTGAGAAAGAAACCAACCACTGCAAGGAATG 1904  
DB 1847 CTGCACCTCTGTCTCTGCTGCTGTTACTATATATTGACCGGAGATTTACAGGAACCTGCCACTCCTG 1906  
  
QY 1905 TCCACCTGACACCTTACCTGTCCATACATCAGGTTCTATGGCAAGAGGGCTTGTATTCCATG 1964

Db 1907 CCCCCCTAACACAAATTCGTGAAAGCCCACCAGCCCTATGGTGTCCAGGCCCTGTGTGCCCTG 1966

QY 1965 CGGGCCTGGGAGTAAACAAATCAGGACCATTCCGGTTTGCTATAGTGACTGCTTTTCTA 2024

Db 1967 TGGTCCAGGACCAAGAACAAAGATCCACTCTCTGTGCTACAATGATTGCACCTTCTC 2026

QY 2025 CCATGAAAAAGAAATCAGATTTTGCACACTATGACTTTAGCAACCTCAGCAGTGTGGGCTC 2084

Db 2027 ACGCAACACTCCAACACGAGACTTTCAACTACAACCTCTCCGCTTTGGCAACACCGTCA 2086

QY 2085 ATTAATGAATGGCCCCAGCTTCACCTCCAAGGAAACAAAATACTTCCATTTCCTTCAATAT 2144

Db 2087 TCTTGCTGGAGGGCCAAGCTTCACCTCCAAGGGTTGAAATACTTCCATCACTTTACCCT 2146

QY 2145 CAGTTTATGTGGCATAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTT 2204

Db 2147 CAGTCTCTGTGGAACACGAGGTAGGAAAATGTCTGTGTGCACCGACAATGTCACTGACCT 2206

QY 2205 TACAGTAAAGAAATAGTGGCAGGGTCAGATGATTACACAAAATTTGGTAGGGSCATTGT 2264

Db 2207 CCGGATTCTTGAGGGTGAGTCAGGG-----TTCTCCAAATCTATCACAGCCCTACGT 2257

QY 2265 ATGCCAGTCAACAATTTATCCTTCTGAAAGTAAGGGTTTCCGAGCAGCCCTTATCATCACA 2324

Db 2258 CTGCCAGGCAGTCATATCCCCCAGAGGTGACAGGCTACAAGGCCGGGTTTCTCTCACA 2317

QY 2325 ATCCATCATTTCTGGCAGATACATTCATAGGAGTCAAGTTGAAACCACATTTGAAAAATAT 2384

Db 2318 GCCTGTGAGCCTTGCTGATCGACTTATTGGGGTGACAACAGATATGACTCTGGATGGAAT 2377

QY 2385 TAATATAAAGAAGATATGTTCCAGTTTCCAACAAGCCAATAACAGATGTGCATTTCTT 2444

Db 2378 CACCTCCCAGCTGAACCTTTTCCACCTGGAGTCCCTTGGGAATACCCGACGTGATCTTCT 2437

QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTTGTATTAAATGCGCGATCAACTGCTGTGAAAAAT 2504

Db 2438 TTATAGGTCCAATGATGTGACCCAGTCCTGCAGTTCTGGGAGATCAACCACCATCCGCGT 2497

QY 2505 GAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTCAGTCCCCCAGCAAGTGCCCAGC 2564

Db 2498 CAGGTGCAGTCCACAGAAAACCTGTCCCTGGAAGTTTGTGTGCTGCCAGGAACGTGCTCAGA 2557

QY 2565 AGGTACCTGTGATGGGTGTACGTTTCTATTTTCCCTGTGGGAGAGTGCTGAAGCTTTGCCCTCT 2624

Db 2558 TGGGACCTGTGATGGCTGCAACTTCCACTTCTGTGGGAGAGCGCGCTGCTTGCCCGCT 2617

QY 2625 GTGTACGGAGCATGACTTCCATGAGATTTGAGGAGCCTGCAAGAGAGAGATTTACAGAAAC 2684

Db 2618 CTGCTCAGTGGCTGACTACCGTGCTATCGTCAGCAGCTGTGTGGTGGGATCCAGAGAC 2677

QY 2685 CTTGTATGTGGAATGAACCTAAATGGTGCAATTAAGGAATTTCTTTGCCCTGAGAAAAA 2744

Db 2678 TACTTACGTGTGGCGAGAACCAAGCTATGCTCTGGTGGCAATTTCTGCTGAGCAGAG 2737

QY 2745 GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGGTGGGAGCCGGTGTGGGAGCTTT 2804

Db 2738 AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCCCTG 2797

QY 2805 TACTGCCGTTTTGCTGGTGGCTCTGACCTGCTACTTCTGGAAGAAAGAAATCAAAAACCTGGA 2864

Db 2798 TACTGCCATCCTGCTCACCCTGTGACCTGCTACTTTTGGAAAAAGAAATCAAAAACCTAGA 2857

QY 2865 ATACAAATATTCCAAGTTAGTAATGACGACTAACTCAAAAGAGAGTGTGAACCTCCCGCTGC 2924

Db 2858 GTACAAGTACTCCAAGCTGGTGATGAATGCTACTCTCAAGGACTGTGACCTGCCAGCAGC 2917

QY 2925 AGACAGTTGTGCTATCATGGAAGGAGAAAGATAATGAAGAGAGAGTTGTATATTTCCAAATA 2984

Db 2918 TGACAGCTGCGCCATCATGGAAGCGGAGGATGTAGAGGAGCAGCTCATCTTTACCAGCAA 2977

QY 2985 ACAGTCACTACTAGGAAAACTCAAAATCTTTTGGCAACCAAGGAAAAAGAACCATTTTGA 3044

Db 2978 GAAGTCACTCTTTGGGAAGATCAAAATCATTTACCTCCAAGAGGACTCCTGATGGATTGA 3037

QY 3045 ATCTGTTCAACTGAAAAACCTC 3065

Db 3038 CTCAGTGCCGCTGAAGACATC 3058

RESULT 7

AAS21262

ID AAS21262 standard; cDNA; 3501 BP.

XX AAS21262;

AC AAS21262;

XX 24-OCT-2001 (first entry)

XX Human cDNA sequence encoding for PRO4985 polypeptide.

DE Human secretory and transmembrane; PRO; mammalian; cancer; lung;

XX breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;

KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;

KW adipocyte; A-peptide; factor VIIA; gene therapy; ss.

XX Homo sapiens.

OS

XX WO200140466-A2.

PN

XX 07-JUN-2001.

PD

XX 01-DEC-2000; 2000WO-US32678.

PF

XX 01-DEC-1999; 99WO-US28301.

XX 01-DEC-1999; 99WO-US28634.

PR 02-DEC-1999; 99WO-US28551.

PR 02-DEC-1999; 99WO-US28564.

PR 02-DEC-1999; 99WO-US28565.

PR 09-DEC-1999; 99US-0170262.

PR 16-DEC-1999; 99WO-US30095.

PR 20-DEC-1999; 99WO-US30911.

PR 20-DEC-1999; 99WO-US30999.

PR 30-DEC-1999; 99WO-US31243.

PR 06-JAN-2000; 2000WO-US00277.

PR 06-JAN-2000; 2000WO-US00376.

PR 11-FEB-2000; 2000WO-US03565.

PR 18-FEB-2000; 2000WO-US04341.

PR 18-FEB-2000; 2000WO-US04342.

PR 22-FEB-2000; 2000WO-US04414.

PR 24-FEB-2000; 2000WO-US04914.

PR 24-FEB-2000; 2000WO-US05004.

PR 01-MAR-2000; 2000WO-US05601.

PR 20-MAR-2000; 2000WO-US07377.

PR 21-MAR-2000; 2000WO-US07532.

PR 30-MAR-2000; 2000WO-US08439.

PR 17-MAY-2000; 2000WO-US13705.

PR 22-MAY-2000; 2000WO-US14042.

PR 30-MAY-2000; 2000WO-US14941.

PR 02-JUN-2000; 2000WO-US15264.

PR 10-NOV-2000; 2000WO-US30873.

(GETH ) GENENTECH INC.

Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;  
Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;  
Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;

WPI; 2001-408281/43.  
P-PSDB; AAU12190.

Isolated, secretory and transmembrane PRO polypeptide used to detect  
other PRO polypeptides, link bioactive molecules to cells expressing  
PRO polypeptides, and detect the presence of mammalian tumours e.g.  
lung, breast, prostate, cervical

Claim 3: Fig 37: 813pp; English.





Db	1995	CCCCCTAACACAAATCTTGAAAGCCCAACGACCTTATGGTGTCCAGGCCGTGTGTGCCCTG	2050
QY	1965	CGGGCCTGGGAGTAAAAACAATCAGGACCAATTCGGTTTGCTATAGTGACTGTCTTTTCTA	2024
Db	2055	TGGTCCAGGGACCAAGAACAACAAGATCCACTCTCTGTGCTACAATGATTGCACCTTCTC	2114
QY	2025	CCATGAAAAAGAAAAATCAGATTTTGCACATGACTTTAGCAACCTCAGCAGTGTGGGCTC	2084
Db	2115	ACGCAACACTCCAACCAAGGACTTTCAACTACAACTTCTCCGCTTTGGCAACACCGCTCAC	2174
QY	2085	ATTAATGAATGGCCCCAGCTTCACCTCCAAGGAACAAAATACTTCCAATTCTTCAATAT	2144
Db	2175	TCTTGTGGAGGCCAAGCTTCACTTCCAAAGGGTTGAAATACTTCCATCAGCTTACCCT	2234
QY	2145	CAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTT	2204
Db	2235	CAGTCTCTGTGGAACCAAGGGTAGGAAAATGTCTGTGCACCCGACAATGTCTCAGCT	2294
QY	2205	TACAGTAAAAGAAATAGTGGCAGGTCAGATGATTACACAAAATTTGGTAGGGGCATTTGT	2264
Db	2295	CCGGATTCCCTGAGGGTGAGTCAGGG-----TTCTCCAAATCTATCACAGCCTACGT	2345
QY	2265	ATGCCAGTCAACAATTATTCCCTTCTGAAAGTAAGGGTTTCCGAGCAGCCTTATCATCACA	2324
Db	2346	CTGCCAGGCAGTCATCATCCCCCAGAGGTTGACAGGCTACAAGGCCGGGTTTCCTCACA	2405
QY	2325	ATCCATCATCTGGCAGATACATTCATAGGAGTCACAGTTGAAACCACATTTGAAAAATAT	2384
Db	2406	GCCTGTGAGCCTTGCTGATCGACTTATTGGGGTGACAACAGATATGACTCTGGATGGAAT	2465
QY	2385	TAATATAAAGAAAGATATGTTCCCAGTTTCCAAACAGCCAAATACCAGATGTGCATTTCCT	2444
Db	2466	CACCTCCCCAGCTGAACCTTTTCCACCTGGAGTCTTGGGAATACCGGACGTGATCTTCTT	2525
QY	2445	TTATAAGTCTCTACAGCAACAACATCTTGTATTAATGCGCCGATCAACTGCTGTGAAAAAT	2504
Db	2526	TTATAGTCCAATGATGTGACCCAGTCCCTGCACTTCTGGGAGATCAACCACCATCCGCGT	2585
QY	2505	GAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGATTTCACTCCCAAGCAAGTGCACG	2564
Db	2586	CAGGTGCAGTCCACAGAAAACTGTCCCTGGAGTTGCTGCTGCCAGGAACGTGCTCAGA	2645
QY	2565	AGGTACCTGTGATGGGTGACGTTCTATTTCTCTGGGAGAGTGTGAAGCTTGCCCTCT	2624
Db	2646	TGGGACCTGTGATGGCTGCAACTTCCACTTCTCTGTGGAGAGCGCGGCTGCTGCCCGCT	2705
QY	2625	GTGTACGGAGCATGACTTCCATGAGATTTGAGGGAGCCTGCAAGAGAGGATTTCAAGAAAC	2684
Db	2706	CTGCTCAGTGGCTGACTACCATGCTATCGTACAGCAGCTGTGTGGTGGGATCCAGANGAC	2765
QY	2685	CTTGATGTGTGAATGAACCTAAATGGTGCAATTAAGGAATTTCTTTGCCCTGAGAAAAA	2744
Db	2766	TACTTACGTGTGNCGAGAAACCAAGCTATGCTCTGGTGGCATTTCTCTGCCCTGAGCAGAG	2825
QY	2745	GTTGGCAACCTGTGAAACCGTTTGACTTTTGGCTGAAGGTGGGAGCCGGTGTGGGAGCTTT	2804
Db	2826	AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG	2885
QY	2805	TACTGCCGTTTTTGGTGGTGGCTCTGACCTGCTACTTCTGGA AAAAAGAATCAAAAACCTGGA	2864
Db	2886	TACTGCCATCCTGCTCACCGTCTTGACCTGCTACTTTTGGAAAAAGAATCAAAAACCTAGA	2945
QY	2865	ATACAAATATTCAGTTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCCGCTGC	2924
Db	2946	GTACAAGTACTCCAAGCTGGTGATGAATGCTACTCTCAAGGACTGTGACCTGCCAGCAGC	3005
QY	2925	AGACAGTTGTGCTATCATGGAAGGAGAAGATAATGAAGAGGAAGTTGTATATTCCAATAA	2984
Db	3006	TGACAGCTGCGCCCATCATGGAAGCGGAGGATGTAGAGGACGACCTCATCTTTACCAGCAA	3065
QY	2985	ACAGTCACTACTAGGAAAACCTCAAATCTTTTGGCAACCAAGGAAAAAGAACCATTTTGA	3044
Db	3066	GAAGTCACTTTTGGGAAGATCAAATCATTTACTCCAAGAGGACTCCTGATGGATTTGA	3125

QY	3045	ATCTGTTCAACTGAAACCTC	3065
Db	3126	CTCAGTGCCGCTGAAGACATC	3146

RESULT 8  
AAF28030  
ID AAF28030 standard; DNA; 3334 BP.  
XX  
AC AAF28030;  
XX  
DT 08-MAY-2001 (first entry)

PS Claim 4; Page 394-398; 418pp; English.

The present invention provides the protein and coding sequences of the human tumour necrosis factor receptors TR13 and TR14. These sequences are useful in the diagnosis and treatment of many diseases, including cancer, autoimmune diseases, cardiovascular disorders, allergies, neurodegenerative diseases, graft rejection, inflammation, aneurysms and infections.

Sequence 3334 BP; 820 A; 952 C; 811 G; 751 T; 0 other;

```
Query Match          27.5%; Score 976.2; DB 22; Length 3334;
Best Local Similarity 60.5%; Pred. NO. 6.1e-246;
Matches 1755; Conservative 0; Mismatches 1118; Indels 28; Gaps 8;
```

QY	171	GCTTCCTCCTTGCCAGGAGAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGG	230
Db	192	GCTTCACGCCTGCAAGAGTCTGAGTACCACATATGAGTACACGGCGTGTGACAGCACGGG	251
QY	231	CTCCAGGTGGAGAGTTGCCATTCCAAATTCGCACTGGACTGCTCTGGCCCTGCCTGACCC	290
Db	252	TTCCAGGTGGAGGGTCGCCGTGCCGATACCCCGGCCCTGTGCACCCAGCCTGCCTGACCC	311
QY	291	AGTGAGAGGCCAAAGAATGCACTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAA	350
Db	312	CGTCAAGGGCACCGAGTGCTCCTTCTCTGCAACGCCGGGGAGTTTCTGGATATGAAGGA	371
QY	351	CCAGGTATGCAGTAAGTGTGGTGAAGGCACCTATTCTTGGGCAGTGGCATCAAAATTGA	410



Db 372 CCAGTCATGTAAGCCATGCGCTGAGGGCCGCTACTCCCTCGGCACAGGCATTCGGTTTGA 431

QY 411 TGAATGGGATGAATTCGGGCAGGATTTTCTAACATCGCAACATTCATGAGCACTGTGGT 470

Db 432 TGAGTGGGATGAGCTGCCCATGGCTTTGCCAGCCTCTCAGCCAAACATGGAGCTGGATGA 491

QY 471 GGGCCCTTCTGACAGCAGGCCAGACGGCTGTAAACACTCTTCTTGGATCCCTCGTGGAAA 530

Db 492 CAGTGTCTGAGTCCA--CCGGAACTGTACTTCGTCCAAGTGGGTTCGCCGGGGCGA 548

QY 531 CTACATAGAATCTAATCGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAA 590

Db 549 CTACATCGCCTTCAACACGGGACGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 608

QY 591 GAAGTCAGGCTATGTCTTCTTGTAGTACCAGTATGTGCACAACAACATCTTCTTTGAGTT 650

Db 609 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 668

QY 651 CTTTATTCAAAATGATCGTCCAGGAGTGGACACCACCACACTGACAAAGTGGGTAAAACT 710

Db 669 TTTCGTTCAGAATGACCAGTCCAGGCCCAATGCAGATGACTC--CAGGTGGATGAAGAC 725

QY 711 TACAGACAATGGAGAAATGGGGCTCTCATTTCTGTAATGCTGAAATCAGGCAACAACTACT 770

Db 726 CACAGAGAA--AGATGGGAATTCACAGTGTGGAGCTAAATCGAGGCAATAATGTCT 782

QY 771 CTACTGGAGAACTACAGGCATCCTTATGGGTCTTAAGCGGTCAAGCCTGTGCTGGTAAA 830

Db 783 CTATTGGAGAACCAAGCCTTCTCAGTATGGACAAAGTACCCAAGCCTGTGCTGGTGAG 842

QY 831 AAATATCACAAATTGAAGGGGTGGCGTACACATCAGAAATGTTTTCCTTGAAGCCAGGCAC 890

Db 843 AAACATTGCCATACAGGGGTGGCCTACACTTCAGAAATGCTTCCCTGCAAACTTGGCAC 902

QY 891 ATTACGAACAACACAGGTTCAATTCAACTGCCAGGTGTGTCGCCAGAAACACCTATTCTGA 950

Db 903 GTATGCAGACAAGCGGCTCCTCTTCTGCAAACTTTGCCAGCCCACTCTTATTCAA 962

QY 951 GAAAGGAGCCAAAGAAATGTATAAGGTGTAA--AGACGACTCTCAATTTTCAGGATCCAG 1007

Db 963 TAAAGGAGAAACTTCTTGCCACCAGTGTGACCTTGACAAATACTCAGAGRAAGGATCTTC 1022

QY 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCACARAAGACTATTTCCAGATCCATCTCCATG 1067

Db 1023 TTCTGTAACTGCGCCAGCTTGACACAGACAAAGATTATTTCTACACACACAGCGGCTG 1082

QY 1068 TGATGAAGAAAGGAAGACACAGATAATGTACAAAGTGGATAGAGCCCAAAATCTGCCGGGA 1127

Db 1083 CGATGCCAACGGAGAGACACAACATCATGTACAAATGGGCCAAAGCCGATACTGTAGCGA 1142

QY 1128 GGATCTCACAGATGCTATTAGATTGCCCCCTTCTTGAGAGAGAAGGATGTGCCGCTTG 1187

Db 1143 GGACCTTGAGGGGGCAGTGAAGCTGCCTGCCTCTGCTGTGAAGACCCACTGCCCACTCTG 1202

QY 1188 CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTGGAACATT 1247

Db 1203 CAACCCAGGCTTCTCAAAACCAACAACAGCACCTGCCAGCCCTGCCCATATGGTTCTTA 1262

QY 1248 TTCAGATGGAACCAAGAAATGTAGACCATGTCCAGCAGGAACGGAGCCCTGCACCTTGGCTT 1307

Db 1263 CTCCAATGG--CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGATT 1319

QY 1308 TGAATATAAATGGTGGAAATGTCTTCTTGGCAACATGAAAACCTTCTGCTCAATGTGG 1367

Db 1320 TGAATACAAATGGTGGAAACACGCTGCCCAACAACATGGAAACGACCTTCTCAGTGGGAT 1379

QY 1368 GAATTTCAAAGTGGCATGGAATGAATGGTTGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427

Db 1380 CAACTTCAGTACAAGGGCATGACAGGCTGGAGGTGGCTGGTGAATCATATATACACAGC 1439

QY 1428 GGCTGGAGGTTCTGACAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAACC 1487

Db 1440 TGCTGGAGCCTCAGACAATGACTTTCATGATTTCTCACTCTGGTTGTGCCAGGATTTAGACC 1499

QY 1488 ACCAACATCTATGACTGGAGCCAC--GGGTCTGAACTAGGAAGAAATAACATTTGTCTT 1544

Db 1500 TCCGCAGTCCGTGATGGCAGACACAGAGAATAAAGAGGTGGCCAGAATCACATTTGTCTT 1559

QY 1545 TGAGACCCCTCTGTTCAGCTGACTGTGTGTTTGTACTTTCATGGTGGATATTATAGAAAAG 1604

Db 1560 TGAGACCCCTCTGTTCTGTGAACTGTGAGCTCTACTTCATGGTGGGTGTGAATTCAGGAC 1619

QY 1605 TACAAATGTGGTAGAATCGTGGGGTGGAAACCAAAAGAAACAAGCTTACACCCAATATCAT 1664

Db 1620 CAACACTCCTGTGGAGACGTGGAAAGGTTCCAAAGGCCAANACAGTCCATATACCTACATCAT 1679

QY 1665 CTTCAAGAAATGCAACTTTTACATTTACATGGGCATTCAGAGAACTAATCAGGGTCAAGA 1724

Db 1680 TGAGGAGAACACTACACAGAGCTTCACCTGGGCTTCCAGAGGAGCCACTTTTCATGAGGC 1739

QY 1725 TAATAGACGGTTTCATCAANTGACATGGTGAAGATTTATTCTATCACAGCCACTAATGCAGT 1784

Db 1740 AAGCAGGAAGTACACCAATGACGTGGCAAGATCTACTCCATCAATGTACCAATGTATAT 1799

QY 1785 TGATGGGGTGGCGTCCCTCATGCCGTGCCGTGTGCCCTCTGAAACAGTCCGGTTTCATC 1844

Db 1800 GAATGGCGTGGCCTCTACTGCGGTGCCGTGCCCTTAGAAGCCTCTGATGTGGGCTCCTC 1859

QY 1845 GTGTGTCCCTGCCCTCCAGGCCACTACATTTGANAAGAAACCAACCAGTGCAGGAATG 1904

Db 1860 CTGCACCTCTTGTCTGTGCTGCTTACTATATTTAGCCGAGATTCAGGAACCTGCCACTCCTG 1919

QY 1905 TCCACCTGACACCTTACCTGTCCATACATCAGGCTTATGGCAAGAGGGCTGTATTTCCATG 1964

Db 1920 CCCCCCTAACACAATTTCTGANAAGCCACCAGCCTTATGGTGTCCAGGCCCTGTGTGCCCTG 1979

QY 1965 CGGGCCTGGAGTAAAAACAATCAGGACCAATTCGGTTTGTATAGTGACTGCTTTTCTA 2024

Db 1980 TGGTCCAGGGACCAAGAACAAAGATCCACTCTCTGTGTACAAATGATTCACCTTCTC 2039

QY 2025 CCATGAAAAAGAAATCAGATTTTGCACATATGACTTTTAGCAACCTCAGCAGTGTGGGCTC 2084

Db 2040 ACGCAACACTCCAAACAGGACTTTCAACTACAACTTCTCCGCTTTGGCAAAACACCGTCAC 2099

QY 2085 ATTAATGAATGGCCAGCTTCACCTCCAAAGGAACAAATACTTCCATTTCTTCAATAT 2144

Db 2100 TCTTGTGGAGGGCCAAAGCTTCATTTCCAAAGGTTGAAATACTTCCATCACTTTACCCCT 2159

QY 2145 CAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTT 2204

Db 2160 CAGTCTCTGTGGAACACCGGTTAGGAAATGTCTGTGTGCACCCGACAATGTCACTGACCT 2219

QY 2205 TACAGTAAAGAAATAGTGGCAGGGTCAAGTATACACAAATTTGGTAGGGCAATTTGT 2264

Db 2220 CCGGATTCCTGAGGGTGAGTCAGGG-----TTCTCCAATCTATCACAGCCTTACGT 2270

QY 2265 ATGCCAGTCAACAATTTATCTCTGAAAGTAAAGGTTTCCGAGCAGCCTTATCATCACACA 2324

Db 2271 CTGCCAGGCAGTCATCATCCCCCAGAGGTGACAGGCTACAAGGCCGGGTTTCCCTCACA 2330

QY 2325 ATCCATCATTTCTGSCAGATACATTCATAGGAGTCAACAGTTGAAACCCACATTTGAAAAATAT 2384

Db 2331 GCCTGTGAGCCTTGTGTGATCGACTTATTTGGGGTGACAACAGATATGACTCTGGATGGANAT 2390

QY 2385 TAATATAAAGAAAGATATGTTCCCAAGTTCCAACAAGCCCAATACCAGATGTGCATTTCTT 2444

Db 2391 CACCTCCCCAGCTGAACCTTTTCCACCTGGAGTCTCTGGGAATACCCGACCTGATCTTCTT 2450

QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTTGTATTAATATGGCCGATCAACTGCTGTGAAAAAT 2504

Db 2451 TTATAGGTCCAATGATGTGACCCAGTCCCTGCAGTCTCTGGGAGATCAACCCACCATCCGCGT 2510

QY 2505 GAGGTGTAATCCTTAAATCTGGAGCAGGAGTGAATTTTCAAGTCCCCCAGCAAGTGTCCCCAGC 2564

Db 2511 CAGGTGCAGTCCACAGAAAACTGTCCCTGGAAAGTTTGTGCTGCCAGGAACGTGCTCAGACA 2570







FT /tag= a  
FT /product= "tumour expressed protein"  
XX

PN WO200131003-A1.

XX 03-MAY-2001.

PF 30-OCT-2000; 2000WO-FR03032.

XX 29-OCT-1999; 99FR-0013629.

PR (FABR ) FABRE MEDICAMENT SA PIERRE.

XX Delneste Y; Magistrelli G, Jeannin P, Bonnefoy J;

PI WPI; 2001-328651/34.  
XX P-PSDB; AAB83850.

PT New nucleic acid, expressed in tumours and lymphoid tissue is useful for  
XX identifying agents for treating tumours and autoimmune disease

PS Claim 2; Page 56-60; 85pp; French.

XX  
CC The present sequence encodes a human protein expressed in tumour cells.  
CC The polynucleotide is useful for screening cDNA/genomic DNA banks and  
CC for cloning isolated DNA; identifying mutant forms of the gene that  
CC encodes a human protein, where the mutations are associated with  
CC abnormal gene expression, or promoters and regulators of the gene,  
CC particularly for diagnosis; for recombinant expression of the derived  
CC protein; as probes and primers for detection and amplification; and  
CC as antisense therapeutics. The tumour expressed protein is useful for  
CC raising specific antibodies and to screen agents that modulate its  
CC activity, bind to it or interact with it. These agents are potentially  
CC useful for treatment or prevention of diseases associated with abnormal  
CC expression/activity of the protein, particularly immunological diseases  
CC (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic  
CC infections.

XX  
SQ Sequence 2733 BP; 656 A; 783 C; 687 G; 607 T; 0 other;  
Query Match 24.2%; Score 860.8; DB 22; Length 2733;  
Best Local Similarity 60.0%; Pred. No. 1.3e-215;  
Matches 1571; Conservative 0; Mismatches 1022; Indels 27; Gaps 7;

QY 171 GCTTCCTCCTTGCCAGGAGAAAGATTATCATCTTTGAATATACGGAATGTGATAGCAGTGG 230

Db 135 GCTTCATGCCTGCAAAGAGTCTGAGTACCACCTATGAGTACACGGCGTGTGACAGCACGGG 194

QY 231 CTCAGGTGGAGAGTTGCCATTCCAAATTTCTGCAGTGGACTGCTCTGGCCTGCCTGACCC 290

Db 195 TTCCAGGTGGAGGTGCGCGTGCGGCATACCCCGGCCCTGTGCACCAGCCTGCCTGACCC 254

QY 291 AGTGAGAGGCAAAGAANTGCATTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAA 350

Db 255 CGTCAAGGGCACCGAGTGTCTCCTGCAACGCCGGGAGTTTCTGGATATGAAGGA 314

QY 351 CCAGGTATGCAGTAAGTGTGGTGAAGGCACCTATTCCCTTGGCCAGTGGCATCAAATTTGA 410

Db 315 CCAGTCATGTAAGCCATGCGCTGAGGGCCGCTACTCCCTCGGCACAGGCATTCGGTTTGA 374

QY 411 TGAATGGGATGAATTGCCGGCAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGT 470

Db 375 TGAGTGGGATGAGCTGCCCATGGCTTTGCCAGCCTCTCAGCCAACATGGAGCTGGATGA 434

QY 471 GGGCCTTCTGACAGCAGGCCAGCGGCTGTAACTCTTCTTGGATCCCTCGTGGAAA 530

Db 435 CAGTGTGCTGAGTCCA---CCGGAACTGTACTTCGTCCAAGTGGGTTCCCGGGGGCGA 491

QY 531 CTACATAGAATCTAATCGTGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAA 590

Db 492 CTACATCGCCTCCAACACGGACGAATGCACAGCCACACTGATGTACGCCGTCAACCTGAA 551

QY 591 GAAGTCAGGCTATGTCTTCTTTGAGTACCAGTATGTGACACAACATCTTCTTTGAGTT 650

Db 552 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGACTCATCTTTGAGTT 611

QY 651 CTTTATTCAAAATGATCAGTGCCAGGAGATGGACACACCACCTGACAAGTGGGTAAAACT 710

Db 612 TTTTCGTTCAGAATGACCAGTGCCAGGCCCAATGCAGATGACTC--CAGGTGGATGAAGAC 668

QY 711 TACAGACAATGGAGAATGGGGCTTCATCTGTAAATGCTGAAATCAGGCACAAAACATACT 770

Db 669 CACAGAGAA--AGGATGGGAATTCACAGTGTGGAGCTAAATCGAGGCAATAATGTCCT 725

QY 771 CTACTGGAGAACTACAGGCATCCTTATGGGTTCTAAGCGGTCAAGCCTGTGCTGTTAAA 830

Db 726 CTATTGGAGAAACACACAGCCTTCTCAGTATGGACCAAAAGTACCCCAAGCCTGTGTTGAG 785

QY 831 AAATATCACAAATTGAAGGGTGGGTGACACATCAGAATGTTTCTTGCAGGCCAGGCAC 890

Db 786 AAACATTGCCATAACAGGGTGGCCTACACTTCAGAATGCTTCCCTGCAAAACCTGGCAC 845

QY 891 ATTACAGCAACAAACCAGGTTTCATTCAACTGCCAGGTGTGCCAGAAACACCTATTCTGA 950

Db 846 GTATGCAGACAAGCAGGGCTCCTCTTCTGCAAACTTTGCCCAAGCCCACTCTTATTCAA 905

QY 951 GAAAGGACCCAAAGAATGTATAAGGTGTAA--AGACGACTCTCAATTTTCAGGATCCAG 1007

Db 906 TAAAGGAGAAACTTCTTGCCACCAGTGTGACCCCTGACAAATACTCAGAGAAAGGATCTTC 965

QY 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCATATCCATG 1067

Db 966 TTCTCTGTAACTGCGCCAGCTTGACAGACAAAAGATTATTTCTACACACACACGGCCTG 1025

QY 1068 TGATGAAGAAGGAAAGACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGA 1127

Db 1026 CGATGCCAACGGGAGACACAACTCATGTACAATGGGCCAAGCCGAAATCTGTAGCGA 1085

QY 1128 GGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAAGAGGATTGTCCGCCTTG 1187

Db 1086 GGACCTTGAGGGGGCAGTGAAGCTGCCTGCCTCTGGTGTGAAGACCCCACTGCCACCC 1145

QY 1188 CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTCTGGAACTT 1247

Db 1146 CAACCCAGGCTTCTTCAAAACCAACAACACAGCACCTGCCAGCCCTGCCCATATGGTCC 1205

QY 1248 TTCAGATGGAACCAAGAATGTAGACCATGTCCAGCAGGAAACGGAGCCTGCACCTTGGCT 1307

Db 1206 CTCCAATGG--CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACCTGCTGTGGGAT 1262

QY 1308 TGAATATAAATGGTGGAAATGTCTTCTTGCAACATGAAAACCTTCTGCTTCAATGTTGG 1367

Db 1263 TGAATACAAATGGTGGAAACACGCTGCCACAAAACATGAAAACGACCGTTCTCAGTGG 1322

QY 1368 GAATTCAAAGTCGGATGGAATGAATGGTGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427

Db 1323 CAACCTTCGAGTACAAGGGCATGACAGSGTGGGAGGTGGTGGTGTACACATTACACAGC 1382

QY 1428 GGCTGGAGGTTCTGACAATGATTACCTGATCTTAAACTTGCCATATCCAGGATTTAAACC 1487

Db 1383 TGCTGGAGCCTCAGACAATGACTTTCATGATTCTCACTCTGTTGTGCCAGGATTTAGACC 1442

QY 1488 ACCAACATCTATGACTGGAGCCAC---GGGTTCTGAACTAGGAAGAAATAACATTTGTCTT 1544

Db 1443 TCCGGCAGTCGGTGATGGCAGACACAGAGAATAAAGAGGTGGCCAGAAATCATATTTGTCT 1502

QY 1545 TGAGACCCCTCTGTTTCAGCTGACTGTGTTTGTACTTTCATGGTGGATTAATAAGAAAAG 1604

Db 1503 TGAGACCCCTCTGTTCTGTGAACCTGTGAGCTCTACTTTCATGGTGGTGTGAATTTCTAG 1562

QY 1605 TACAAATGTGGTAGAATCGTGGGTGGAACCAAGAAAGAAAGCTTACACCCATATCAT 1664

Db 1563 CAACACTCCTGTGGAGACGTGGAAAGGTTCCAAAGGCAACAGTCCCTATACCTACATCAT 1622

QY 1665 CTTCAAGAATGCAACTTTTACATTTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAG 1724







DT 17-DEC-2001 (first entry)  
XX Human cDNA encoding a novel foetal antigen, SEQ ID NO 689.  
DE  
XX  
KW Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;  
KW immunomodulator; cardiovascular; cytostatic; nephrothropic;  
KW cardiovascular; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; breast neoplasm; cancer;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; food additive.  
XX  
OS Homo sapiens.  
XX  
PN WO200155312-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01321.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.



PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX  
DR WPI; 2001-488782/53.  
DR P-PSDB; AAU21345.  
XX  
PT New polynucleotides and polypeptides for diagnosing, treating,  
PT preventing or prognosing e.g. diseases or disorders of the nervous,  
PT musculoskeletal, excretory, gastrointestinal, reproductive, and  
PT respiratory systems  
XX  
PS Claim 1; SEQ ID No 689; 642pp; English.  
XX  
CC The invention relates to novel nucleic acids encoding novel human foetal  
CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.  
CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. The antibodies to the antigens can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. Numerous  
CC examples of diseases and disorders treated by the nucleic acids and  
CC proteins are given in the specification. The present sequence

Query Match 18.1%; Score 645.4; DB 22; Length 672;  
Best Local Similarity 98.4%; Pred. No. 2.7e-159;  
Matches 662; Conservative 9; Mismatches 0; Indels 2; Gaps 2;

QY 2868 CAAATATCCCAAGTTAGTAAATGACGACTAACTCAAAGAGTGTGAACCTCCCGCTGCACA 2927  
Db 2 CAAATATCCCAAGTTAGTAAATGACGACTAACTCAAAGAGTGTGAACCTCCCGCTGCAGA 61  
QY 2928 CAGTTGTGCTATCATGGAAGGAGAGATAATGAAGAGGAAGTTGTATATTTCCAATAAACA 2987  
Db 62 CAGTTGTGCTATCATGGAAGGAGAGATAATGAAGAGGAAGTTGTATATTCMNATAAACM 121  
QY 2988 GTCACACTAGGAAAACCTCAAGATCCCCAAATATATGAAGAGACAGTGTGTAGCTTGAG 3047  
Db 122 GTCACACTAGGAAAACCTCAAGATCCCCAAATATATGAAGAGACAG-KCTGTAGCCTTGAG 181  
QY 3048 TGTCAACTGAAAACCTCAAGATCCCCAAATATATGAAGAGACAGTGTGTAGCTTGAG 3107  
Db 182 TGTCAACTGAAAACCTCAAGATCCCCAAATATATGAAGAGACAG-KCTGTAGCCTTGAG 240  
QY 3108 ACTAATGAACAAAGAACCTGCTCTAGTTTACAGGACCATATTTTAGGGTCTGTCTCTCA 3167  
|||||

Db 241 ACTAATGAACAAACCTGCTCTAGTTTACAGGACCATATTTAGGGTCTGTCTCTCA 300  
QY 3168 TACCTGTACATTGGTGATCTCACAGAGGAGGGCCCATGCCGCTGAAAAGGGAAGGAGATT 3227  
|||||  
Db 301 TACCTG-CACATTGGTGATCTCACAGAGGAGGGCCCATGCCGCTGAAAAGGGAAGGAGATT 359  
QY 3228 GAAACATTTGATTGCCCTTATCACATGGTCAAGTACCTTGCCAAATAAAGGAAAGCAAATG 3287  
|||||  
Db 360 GAAACATTTGATTGCCCTTATCACATGGTCAAGTACCTTGCCAAATAAAGGAAAGCAAATG 419  
QY 3288 ATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAAGAGATTATCTGTATATACACA 3347  
|||||  
Db 420 ATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAAGAGATTATCTGTATATACACA 479  
QY 3348 TAACTGAAAACCAAAGTTTAAAGCCACCAAATGCACCTGCTGATGCATGCCATATAATTAATG 3407  
|||||  
Db 480 TAACTGAAAACCAAAGTTTAAAGCCACCAAATGCACCTGCTGATGCATGCCATATAATTAATG 539  
QY 3408 GGTAACCTTTATTTATTTATGATGTCTACATAACAAGTGTGATTGGAAAGGCACATGTGAG 3467  
|||||  
Db 540 GGTAACCTTTATTTATTTATGATGTCTACATAACAAGTGTGATTGGAAAGGCACATGTGAG 599  
QY 3468 CATATGCATTTATGATCCAAATTTATGTTTTTTCTTTTGTATATTTGGGGAAAAATAAAA 3527  
|||||  
Db 600 CATATGCATTTATGATCCAAATTTATGTTTTTTCTTTTGTATATTTGGGGAAAAATAAAA 659  
QY 3528 TTTTNTTTTAAAGGTA 3540  
|||||  
Db 660 TTTTNTTTTAAAGGTA 672

RESULT 13  
ABL69028/c  
ID ABL69028 standard; DNA; 576 BP.  
XX  
AC ABL69028;  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DE Kidney cancer related gene sequence SEQ ID NO:7365.  
XX  
KW Human; cancer; colon; breast; kidney; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200194629-A2.  
XX  
PD 13-DEC-2001.  
XX  
PF 30-MAY-2001; 2001WO-US10838.  
XX  
PR 05-JUN-2000; 2000US-209473P.  
PR 05-JUN-2000; 2000US-209531P.  
PR 18-SEP-2000; 2000US-233133P.  
PR 18-SEP-2000; 2000US-233617P.  
PR 20-SEP-2000; 2000US-234009P.  
PR 20-SEP-2000; 2000US-234034P.  
PR 20-SEP-2000; 2000US-234052P.  
PR 22-SEP-2000; 2000US-234509P.  
PR 22-SEP-2000; 2000US-234567P.  
PR 25-SEP-2000; 2000US-234923P.  
PR 25-SEP-2000; 2000US-234924P.  
PR 25-SEP-2000; 2000US-235077P.  
PR 25-SEP-2000; 2000US-235082P.  
PR 25-SEP-2000; 2000US-235134P.  
PR 25-SEP-2000; 2000US-235280P.  
PR 26-SEP-2000; 2000US-235637P.  
PR 26-SEP-2000; 2000US-235638P.  
PR 27-SEP-2000; 2000US-235711P.  
PR 27-SEP-2000; 2000US-235720P.



PR 18-AUG-1997; 97US-0055684.  
PR 18-AUG-1997; 97US-0055722.  
PR 18-AUG-1997; 97US-0055723.  
PR 18-AUG-1997; 97US-0055947.  
PR 18-AUG-1997; 97US-0055948.  
PR 18-AUG-1997; 97US-0055949.  
PR 18-AUG-1997; 97US-0055950.  
PR 18-AUG-1997; 97US-0055953.  
PR 18-AUG-1997; 97US-0055954.  
PR 18-AUG-1997; 97US-0055964.  
PR 18-AUG-1997; 97US-0055984.  
PR 18-AUG-1997; 97US-0056360.  
PR 12-SEP-1997; 97US-0058660.  
PR 12-SEP-1997; 97US-0058661.  
PR 12-SEP-1997; 97US-0058664.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Brewer LA, Ebner R, Fischer CL, Kyaw H, Lafleur DW, Li Y, Moore PA;  
PI Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR, Zeng Z;  
XX  
DR WPI; 1999-120770/10.  
DR P-PSDB; AAY02679.  
XX  
PT New isolated human genes and the secreted polypeptides they encode -  
PT useful for diagnosis and treatment of e.g. cancers, neurological  
PT disorders, immune diseases, inflammation or blood disorders  
XX  
PS Claim 1; Page 262; 464pp; English.  
XX  
CC This sequence represents a nucleic acid molecule which encodes a  
CC secreted human protein. The gene number, and the clone it is derived  
CC from, are detailed in the descriptor line. The gene can be used to  
CC generate fusion proteins by linking to the gene to a human immunoglobulin  
CC Fc portion (e.g. AAX27302) for increasing the stability of the fused  
CC protein as compared to the human protein only.  
CC The invention relates to 123 novel genes and their fragments (nucleic  
CC acid sequences: AAX27311-X27449; amino acid sequences AAY02650-Y02788)  
CC which are useful for preventing, treating or ameliorating medical  
CC conditions e.g. by protein or gene therapy. Also, pathological  
CC conditions can be diagnosed by determining the amount of the new  
CC polypeptides in a sample or by determining the presence of mutations in  
CC the new polynucleotides. Specific uses are described for each of the 123  
CC polynucleotides, based on which tissues they are most highly expressed in  
CC (see AAX27311 for described uses).  
XX  
SQ Sequence 1129 BP; 379 A; 194 C; 192 G; 363 T; 1 other;

Query Match 13.9%; Score 494; DB 20; Length 1129;  
Best Local Similarity 97.9%; Pred. No. 2.6e-119;  
Matches 511; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 3022 AAGGAAAAGAGACCATTTTGAATCTGTTCAACTGAAAACCTCAAGATCCCCAAATATA 3081  
| | | | |  
Db 442 AGGAAAAGAGAGACCCATTTTGAATCTGTTCAACTGAAAACCTCAAGATCCCCAAATATA 501  
| | | | |  
QY 3082 TGAAGAGACAGTGTCTGTAG-CCTTGAGACTAATGAACAAAGAAACCTGCTCTAGTTTAC 3140  
| | | | |  
Db 502 TGAAGAGACAGTGTCTGTAGCCCTTGAGACTAATGAACAAAGAAACCTGCTCTAGTTTAC 561  
| | | | |  
QY 3141 AGGACCATATTTTAGGGTCTGTCCTCATACCTGTGCATTTGGTGATCTCACAGAGGAGGG 3200  
| | | | |  
Db 562 AGGACCATATTTTAGGGTCTGTCCTCATACCTGTGCATTTGGTGATCTCACAGAGGAGGG 621  
| | | | |  
QY 3201 CCATGCCCTGAAAAGGGAAGGAGATTGAAACATTTTGATTCCTTTATCACATGGTCAAGT 3260  
| | | | |  
Db 622 CCATGCCCTGAAAAGGGAAGGAGATTGAAACATTTTGATTCCTTTATCACATGGTCAAGT 681  
| | | | |  
QY 3261 ACCTTGCCAAATAAGGAAAGCAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTC 3320  
| | | | |  
Db 682 ACCTTGCCAAATAAGGAAAGCAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTC 741  
| | | | |  
QY 3321 AGGAAGAGATTTATCTGTATATACACATAAAGTGAACCAAGTTTAAGCCCAATGCA 3380  
| | | | |

Db 742 AGGAAGAGATTTATCTGTATATACACATAAAGTGAACCAAGTTTAAGCCCAATGCA 801  
| | | | |  
QY 3381 CTGCTGATGCATGCCATATATAATTAATGGGTAACCTTTTATTTATGATGCTACATAAC 3440  
| | | | |  
Db 802 CTGCTGATGCATGCCATATATAATTAATGGGTAACCTTTGATTTTATGACGCTACATAAC 861  
| | | | |  
QY 3441 AAGTGTGATTTTGAAGGCACATGTGAGCATATGCATTATGATCCAAATTTATGTTTTTCT 3500  
| | | | |  
Db 862 AAGTGTGATTTTGAAGGCACATGTGAGCATATGCATTATGATCCAAATTTATGTTTTTCT 921  
| | | | |  
QY 3501 TTGTTTATATTTTGGGGAAAATTAATAATTTTAAAGGTAAA 3542  
| | | | |  
Db 922 TTGTTTATATTTTGGGGAAAATTAATAATTTTAAAGGTATA 963  
| | | | |

RESULT 15  
AAZ41991  
ID AAZ41991 standard; cDNA; 1717 BP.  
XX  
AC AAZ41991;  
XX  
DT 31-JAN-2000 (first entry)  
XX  
DE Human endometrium tumour cDNA derived EST 11.  
XX  
KW Endometrium; human; tumour; cancer; anticancer; cytostatic; EST:  
KW treatment; uterine; gene therapy; expressed sequence tag; ss.  
XX  
OS Homo sapiens.  
XX  
PN DE19817948-A1.  
XX  
PD 21-OCT-1999.  
XX  
PF 17-APR-1998; 98DE-1017948.  
XX  
PR 17-APR-1998; 98DE-1017948.  
XX  
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.  
XX  
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E;  
XX  
DR WPI; 1999-591957/51.  
DR P-PSDB; AAY59971, AAY59972, AAY59973.  
XX  
PT New nucleic acid sequences expressed in uterine cancer tissues, and  
PT derived polypeptides, for treatment of uterine and endometrial cancer  
PT and identification of therapeutic agents -  
XX  
PS Claim 3; Page 172; 444pp; German.  
XX  
CC This invention describes novel human nucleic acid (cDNA) sequences (A),  
CC that are highly expressed in uterine tumour tissue and which have  
CC anticancer and cytostatic activity. (A) are used (i) for recombinant  
CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)  
CC are used (i) to identify agents suitable for treatment of uterine or  
CC endometrial cancer; (ii) directly for treating these forms of cancer  
CC (including expression from gene therapy vectors) and (iii) for  
CC generation of specific antibodies. (A) are identified by assembling ESTs  
CC (expressed sequence tags) from a particular tissue type before comparison  
CC of expression patterns. This allows a significantly longer fragment of  
CC the gene to be revealed, so should reduce the number of failures  
CC associated with the fact that ESTs from different libraries may represent  
CC different parts of the same unknown gene, distorting the estimated  
CC frequency of occurrence in a particular tissue. AAZ41981-242121 represent  
CC EST fragments derived from a human endometrium tumour cDNA library which  
CC encode the protein sequences represented in AAY59941-Y60328.  
XX  
SQ Sequence 1717 BP; 416 A; 492 C; 388 G; 421 T; 0 other;

Query Match 13.7%; Score 487; DB 20; Length 1717;  
Best Local Similarity 59.1%; Pred. No. 2.3e-117;



Matches 857;		Conservative	0;	Mismatches	585;	Indels	9;	Gaps	1;
QY	1615	GTAGAATCGTGGGGTGGAAACCAAGAAACAAAGCTTACACCCATATCATCTTCAAGAAAT							1674
Db	21	GTGGAGAGCTGGAAAGGTTCCAAAGGCAACAGTCCTATACCTACATCATTTGAGGAGAAC							80
QY	1675	GCAACTTTTACATTTACATGGGCATTCACAGAGAACTAATCAGGGTCAAGATAATAGACGG							1734
Db	81	ACTACCAGAGCTTCACCTGGGCTTCCAGAGGACCACCTTTTCATGAGGCAAGCAGGAAG							140
QY	1735	TTCATCAATGACATGGTGAAGATTTATTTCTATCACAGCCACTAATGCAGTTGATGGGGTG							1794
Db	141	TACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCACCAATGTTATGAATGGCGTG							200
QY	1795	GGTCTCATGCGCTGCCGTGCGCCTCGGTTCTGAACAGTCGGGTTTCATCGTGTCTCCCC							1854
Db	201	GCCTCCTACTGCCGTGCCCTGCCCCTAGAAGCCTCTGATGTGGCTCCTCGCACCTCT							260
QY	1855	TGCCCTCCAGGCCACTACATTTGAGAAAGAAACCAACCAGTGCAGGAATGTCCACCTGAC							1914
Db	261	TGTCCTGCTGGTTACTATATTGACCGAGATTCAGGAACCTGCCACTCCTGCCCCCCCTAAC							320
QY	1915	ACCTACCTGTCCATACATCAGGCTATGGCAAAGAGGCTTGTATTCCATGCGGGCCTGGG							1974
Db	321	ACAATCTGAAAGCCCCACACGCTTATGGTGTCCAGGCTGTGTGCCCTGTGGTCCAGGG							380
QY	1975	AGTAAAAACAATCAGGACCATTCCGTTTGTGTATAGTGACTGCTTTTTCACCATGAAGAAA							2034
Db	381	ACCAAGAACAACAAGATCCACTCTCTGTGCTACAAATGATTGCACCTTCTCAGCAACACT							440
QY	2035	GAAATCAGATTTTGCACATATGACTTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAAT							2094
Db	441	CCAACCAGGACITTTCAACTACAACTTCTCCGCTTGGCAACACCGTCACTCTTGCTGGA							500
QY	2095	GGCCCCAGCTTCACCTCCAAGGAACAATAACTTCCATTTCTTCAATATCAGTTTATGT							2154
Db	501	GGCCAAAGCTTCATTCCAAAGGGTTGAAATACCTCCATCACTTTACCCCTCAGTCTCTGT							560
QY	2155	GGGCATGAGGGGAAGATGGCTCTCTGTACCAACAATAATAACAGACTTTACAGTAAAA							2214
Db	561	GAAACCAGGGTAGGAAATGTCTGTGTGCACCGACAAATGTCACTGACCTCCGGATTCCCT							620
QY	2215	GAAATAGTGGCAGGTCAGATGATTACACAAATTTGGTAGGGCATTTGTATGCCAGTCA							2274
Db	621	GAGGGTGAGTCAGG-----TTCTCCAATCTATCACAGCCTACGCTGCCAGGCA							671
QY	2275	ACAATTATTCCTTCTGAAAGTAAGGTTTCCGAGCAGCCTTATCATCACAAATCCATCAT							2334
Db	672	GTATCATCCCCCGAGAGGTGACAGGCTACAAAGCCGGGTTTCTCTCACAGCCTGTCTCAG							731
QY	2335	CTGGCAGATACATTCATAGGAGTCACAGTTGAAACCACATTTGAAATAATATTATAAAA							2394
Db	732	CTTGCTGATCGACTTATGGGGTGACAACAGATATGACTCTGGATGGAATCACCTCCCCA							791
QY	2395	GAAGATATGTTCCAGTTCCAACAAGCCCAATACCAGATGTGCATTTCTTTTATAAGTCT							2454
Db	792	GCTGAACITTTCCACCTGGAGTCCTTGGGAATACCGGACGTGATCTCTTTTATAGGTCC							851
QY	2455	TCTACAGCAACACATCTTGTATTAAATGGCCGATCAACTGCTGTGAAAAATGAGGTGTAAT							2514
Db	852	AATGATGTGACCCAGTCCTCGCAGTTCTGGGAGATCAACCACCATCCGCGTCAGGTGCAG							911
QY	2515	CCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCCAGCAAGTCCCCCAGCAGGTACCTGT							2574
Db	912	CCACAGAAAACGTGCCCTGGAGTTTGTGCTGCCAGGAACGTGCTCAGATGGGACCTGT							971
QY	2575	GATGGGTGACGTTCTATTCTCTGGGAGAGTGTGAAGCTTGCCCTCTGTGTACGGAG							2634
Db	972	GATGGCTGCAACTTCCACTTCTGTGGGAGAGCGCGGCTGCTTGCCCGCTCTGCTCAGTG							1031
QY	2635	CATGACTTCCATGAGATTGAGGGAGCCTGCAAGAGAGGATTTACAGGAACCTTGTATGTG							2694
Db	1032	GCTGACTACCATGCTATCGTCAGCAGCTGTGTGGCTGGGATCCAGAAGACTACTTACGTG							1091

Search completed: May 12, 2003, 00:39:48  
Job time : 738.514 secs

QY	2695	TGGAATGAACCTAAATGGTGCATTAAAGGAATTTCTTTGCCTGAGAAAAAGTTGGCAACC	2754
Db	1092	TGGCGAGAACCCCAAGCTATGCTCTGGTGGCAATTTCTCTGCCTGAGCAGAGTACCATC	1151
QY	2755	TGTGAACCGTTGACTTTTGGCTGAAGGTGGAGCCGGTGTGGGAGCTTTTACTGCCGTT	2814
Db	1152	TGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTGTACTGCCATC	1211
QY	2815	TTGCTGGTGGCTCTGACCTGTACTTCTGGAAGAAAGAAATCAAAAACTGGAATACAAATAT	2874
Db	1212	CTGCTCACCGCTCTTGACCTGTACTTTTGGAAAAAGAAATCAAAAACTAGAGTACAAGTAC	1271
QY	2875	TCCAAGTTAGTAATGACGACTAACTCAAAAGAGTGTGAACTCCCCGGCTGCAGACAGTTGT	2934
Db	1272	TCCAAGCTGGTGATGAATGCTACTCTCAAGGACTGTGACCTGCCAGGAGCTGACAGCTGC	1331
QY	2935	GCTATCATGGAGGAGAGATAATGAAGAGGAAAGTTGTATATTTCCAATAAACAGTCACTA	2994
Db	1332	GCCATCATGGAGGCGGAGGATGTAGAGGACGACCTCATCTTTACCAGCAAGAGTCACTC	1391
QY	2995	CTAGGAAAACTCAAACTCTTTGGCAACCAAGGAAAAAGAGACCAATTTTGAATCTGTTCAA	3054
Db	1392	TTTGGGAAGATCAAAATCATTTACCTCCAAGAGGACTCCTGTATGGATTTGACTCAGTGCCG	1451
QY	3055	CTGAAAAACCTC	3065
Db	1452	CTGAAGACATC	1462

BACKFILE DOCUMENT INDEX SHEET

REV 7/02/03

APPL PARTS

A DOCPHOENIX

Internal Misc. Paper	IMIS
Misc. Incoming Letter	LET.

PCT Papers in a 371Application	371P
Amendment Including Elections	A...
Abstract	ABST
ADS	ADS
Application Data Sheet	AF/D
Affidavit or Exhibit Received	APPENDIX
Artifact	ARTIFACT
Bib Data Sheet	BIB
Claim	CLM
Computer Program Listing	COMPUTER
CRFL	CRFL
All CRF Papers for Backfile	DIST
Terminal Disclaimer Filed	DRW
Drawings	FOR
Foreign Reference	FRPR
Foreign Priority Papers	IDS
IDS Including 1449	

Internal	SRNT	Examiner Search Notes	CLMPTO	PTO Prepared Complete Claim Set
ECBOX	WCLM	Claim Worksheet	WFEE	Fee Worksheet
		Evidence Copy Box Identification		

Misc. Office Action	CTMS
Signed 1449	1449
892	892
Abandonment	ABN
Board of Appeals Decision	APDEC
Examiner Answer	APEA
Count Advisory Action	CTAV
Count Ex parte Quayle	CTEQ
Count Final Rejection	CTFR

OUTGOING

Non-Patent Literature	NPL
Oath or Declaration	OATH
Petition	PET.
Mail Returned by USPS	RETMAL
Sequence Listing	SEQLIST
Specification	SPEC
Specification Not in English	SPEC NO
Transmittal New Application	TRNA

INCOMING

Count Non-Final	CTNF
Count Restriction	CTRS
Examiner Interview	EXIN
DO/EO Acceptance	M903
DO/EO Missing Requirement	M905
Formal Drawing Required	NFDR
Notice of Allowance	NOA
Petition Decision	PETDEC

AP.B	Appeal Brief
C.AD	Change of Address
N/AP	Notice of Appeal
PA..	Change in Power of Attorney
REM	Applicant Remarks in Amendment
XT/	Extension of Time filed separate

File Wrapper

FWCLM	File Wrapper Claim
IIFW	File Wrapper Issue Information
SRFW	File Wrapper Search Info



